AFOSR 68-1403 Vol II Parts A+B

23

AD 670607

This document has been approved for public release and sale; its distribution is unlimited

JUN 27 1968

Science Communication, Inc.

1081 Wisconsin Ave., N.W. Washington, D. C. 20007

433

AD 670607

STUDY OF SCIENTIFIC AND
TECHNICAL DATA ACTIVITIES
IN THE UNITED STATES --

VOLUME II

PRELIMINARY CENSUS OF SCIENTIFIC
AND TECHNICAL DATA ACTIVITIES
Parts A and B

Prepared for

Task Group on National Systems
Committee on Scientific and Technical Information
Federal Council for Science and Technology

DEC 1968

Final Report
Contract F44620-67-C-0022
ARPA Order: 892 as Amended

SCIENCE COMMUNICATION, INC.
1050 31st Street, N.W
Washington, D.C. 20007

Science Communication
Washington, D. C. 200 07
COSATI Data Activities Study
Final Report - F44620-67-C-0022 30 April 1968

SPONSORSHIP ACKNOWLEDGEMENTS

The work reported herein was sponsored by the Task Group on National System(s), Committee on Scientific and Technical Information (COSATI) of the Federal Council for Science and Technology. Financial support was provided by the Advanced Research Projects Agency of the Department of Defense, with contract administration by the Air Force Office of Scientific Research. Work was initiated, under Air Force Contract F44620-67-C-0022, on 1 September 1966 and completed 30 April 1968.

FEDERAL COUNCIL FOR SCIENCE AND TECHNOLOGY
COMMITTEE ON SCIENTIFIC AND TECHNICAL INFORMATION
EXECUTIVE OFFICE BUILDING
WASHINGTON, D.C. 20506

DEC 1968

FOREWORD

The Task Group on National Systems for Scientific and Technical Information of the Committee on Scientific and Technical Information (COSATI) is sponsoring a series of studies on aspects of information systems and activities in the United States. This report by Science Communication, Inc., is the result of one such study.

COSATI feels that this report contains much valuable information and many thought-provoking recommendations. Both government and private communities should benefit by having the report widely distributed, and extensively reviewed and discussed. Hopefully professional societies, private groups and interested individuals will continue the analysis of scientific and technical data activities which has been well begun in this report.

Andrew A. Aines
Chairman

Science Communication
Washington, D. C. 200 07
COSATI Data Activities Study
Final Report - F44620-67-C-0022

30 April 1968

ABSTRACT

This volume presents the findings from a preliminary survey of scientific and technical data activities in industry, the professions, and government. The purpose of the survey was compilation of information which could support the development of national policies and plans with respect to data management and data handling systems. The survey constitutes one of a complementary set of exploratory studies sponsored by the Task Group on National System(s) of the Committee on Scientific and Technical Information (COSATI). COSATI is a committee of the Federal Council for Science and Technology.

The survey scope, roughly defined, includes the more important data activities supporting our national science-technology effort. Emphasis is directed to those data activities and formal data handling efforts which would most likely be considered in conjunction with planning and development of national data systems.

This volume consists of three parts. Part A presents scenarios of data activities in ten selected fields of science or technology. Each scenario covers the characteristics of data, data flows, formal data efforts, and representative data related problems or issues identifiable with the field The fields covered are: aerospace science and technology, electronics and electrical engineering, materials science and engineering, chemistry and chemical engineering, agriculture and food technology, biomedical sciences, pharmacology, social and behavioral sciences, environmental and geosciences, and oceanography. A supplementary scenario describes data activities as conducted within the research, developmental, and applications phases of scientific and technological activity.

Part B summarizes results from probes of selected areas of scientific and technical data activity. Areas probed include data activities of medical research institutions, professional societies and trade associations, commercial data processing service centers, and U. S. Army research, development, test and evaluation activities. Part B also includes a review of equipment capabilities.

Part C consists of a preliminary census of 226 formal data efforts which are representative of those efforts currently operating in the United States. The following types of data efforts are included in the census: Data service centers, Data-document depositories, Data program development and coordination, Non-designated (Agency) data handling and service operations, and Small evolving data handling and service operations,

The information contained in this volume supported the preparation of a plan for actions to improve existing data systems and to further explore the feasibility of national data system concepts. This plan is outlined in Volume I of this report.

TABLE OF CONTENTS

VOLUME II

TABLE OF CONTENTS (Cont'd)

Science Communication
Washington, D. C. 200 07

COSATI Data Systems Study
Final Report - F44620-67-C-0022 30 April 1968

ACCESSIBILITY OF DOCUMENTS CITED IN THIS REPORT

Many of the background documents for this study are reports of Government sponsored studies. Most of these documents are available from the Clearinghouse for Federal Scientific and Technical Information ("CFSTI"), Springfield, Va., 22151. In ordering Clearinghouse documents, use of the "PB" or "AD" numbers is suggested to expedite the processing. The other principal source of government-sponsored documents is the Superintendent of Documents, Government Printing Office ("GPO"), Washington, D.C. 20402.

It is the policy of the President's Science Advisory Committee and the Federal Council for Science and Technology, Committee on Scientific and Technical Information, to make their reports and reports sponsored by them readily available to the public. To assist the reader, therefore, the following information supplements the bibliographic references to such reports as they appear in this report:

1. Progress of the United States Government in Scientific and Technical Communications, Committee on Scientific and Technical Information of the Federal Council for Science and Technology, Executive Office of the President, 1965, PB 173 510. Available from CFSTI.

2. Recommendations for National Document Handling Systems in Science and Technology: Appendix A -- A Background Study -- Volumes I and II, System Development Corporation, Santa Monica, California, September 1965, AD 624 560, PB 168 267. Available from CFSTI.

3. A System Study of Abstracting and Indexing in the United States, System Development Corporation, Falls Church, Virginia, 16 December 1966, PB 174 249. Available from CFSTI.

4. Exploration of Oral/Informal Technical Communications Behavior, Semi-Annual Technical Report, American Institutes for Research, Silver Spring, Maryland, 15 March 1967, AD 650 219. Available from CFSTI.

5. Handling of Toxicological Information, A Report of the President's Science Advisory Committee, The White House, Washington, D.C., June 1966. Available from GPO.

6. Science, Government, and Information, A Report of the President's Science Advisory Committee, The White House, January 10, 1963, GPO (out of print).

7. Report of the Office of Science and Technology Ad Hoc Panel on Scientific and Technical Communications, J.C.R. Licklider, et al., 8 February 1965. Unpublished (out of print).

8. The Copyright Law as it Relates to National Information Systems and National Information Systems and National Programs -- a Study by the COSATI Ad Hoc Task Group on Legal Aspects Involved in National Information Systems, Washington, D.C., July 1967, PB 175 618. Available from CFSTI.

9. Progress of the United States Government in Scientific and Technical Information, Committee on Scientific and Technical Information (COSATI) of the Federal Council for Science and Technology, Washington, D.C., 1966, PB 176 535. Available from CFSTI.

10. Review of Proposal for a National Data Center, Statistical Evaluation Report No. 6, Edgar S. Dunn, Jr., Office of Statistical Standards, Bureau of the Budget, December 1965. Available from Bureau of the Budget, Executive Office Building, Washington, D.C. 20506.

11. President's Message on Communications Policy to the Congress of the United States. The White House, Washington, D.C., August 14, 1967. Available White House Press Office, Washington, D.C., 20506.

12. Scientific and Technological Communication in the Government, (The Crawford Report), Task Force Report to the President's Special Assistant for Science and Technology, Washington, D.C., April 1962, AD 299 545. Available from CFSTI.

13. Information Sciences Technology: First Report of Panel 2, Committee on Scientific and Technical Information of the Federal Council for Science and and Technology, September 1965, PB 169 686. Available from CFSTI.

14. Presidential Message upon signing of the State Technical Services Act, P.L. 89-182, President Lyndon B. Johnson, September 14, 1965. Available from White House Press Office, Washington, D.C., 20506.

INTRODUCTION TO VOLUME II

Content and Objective of the Census

It is axiomatic that the scientific and technical data and information systems of the future should be based on a clear understanding of current activities. Such an understanding can be achieved only through identification of relevant activities, definition of significant elements and characteristics of these activities, and systematic examination of these activities to articulate fundamental structures, functions, and objectives. Consequently, the Committee on Scientific and Technical Information (COSATI) Task Group on National Systems established an objective to inventory and evaluate the resources currently being utilized in national and other selected domestic scientific and technical information and data activities. More specifically, the Task Group has undertaken to:

- Determine why and how the scientist, engineer, manager, and technical public obtain and use scientific and technical information and identify trends that may change these patterns;
- Examine the relationships between generators, processors, users, and systems of scientific and technical data and information to ascertain functions, volumes, economics, trends, problems, etc., both present and future;
- Identify and examine data and information activities being pursued or under development which are of sufficient importance to our national scientific and technical posture to warrant close coordination;
- Consider the development of national data and information systems in relation to trends and requirements as revealed in activities both at the sub-national and international levels; and

- Review the state-of-the-art pertaining to equipments, facilities, techniques, and organizational capabilities as related to existing and potential national data and information system requirements.

The Task Group has sponsored a complementary set of studies to accumulate and articulate background information relevant to its investigation of the requirements and feasibility factors relating to national scientific and technical information system concepts. The first study examined the current status of document handling activities and made recommendations concerning a national document handling system.* A second study dealt in depth with abstracting and indexing services in the United States.** Another study analyzed the structures and functions of informal information-communication systems.*** Reported herein is an exploratory examination of the scientific and technical data activities and related systems currently operational or under development. Emphasis is directed to those data activities, formal efforts, and systems which would most likely be considered in conjunction with planning and development of national systems.

Based upon results of the above studies and other findings, the Task Group is formulating recommendations and plans for the development of national information and data systems which include actions for government agencies, suggestions for actions by the private sector, and steps to move from current to advanced systems. The Task Group is currently considering a plan for actions to improve existing data systems and to further explore the feasibility of national data system concepts. Development of this plan, which is presented in Volume I of this report, was supported by the background information contained in this Volume.

---

* Recommendations for National Document Handling Systems in Science and Technology: Appendix A -- A Background Study -- Volumes I and II, System Development Corporation, Santa Monica, California, September 1965. Contract AF 19 (628) - 5166.

** A System Study of Abstracting and Indexing in the United States, System Development Corporation, Falls Church, Virginia, 16 December 1966. Contract NSF-C-464.

*** Exploration of Oral/Informal Technical Communications Behavior, Semi-Annual Technical Report, American Institutes for Research, Silver Spring, Maryland, 15 March 1967. DAHC-04 67 C0004.

Scope of Census Efforts

In September 1966, Science Communication, Inc. undertook the development of a preliminary census of scientific and technical data activities in industry, the professions, and government. As noted above, the end objective was compilation of information which could support the development of national policy with respect to systems for data collection, reduction, storage, retrieval, analysis, and dissemination. The census scope, roughly defined, was intended to include the more important data activities supporting the national science-technology effort. Specifically, the scope was defined as including data activities involving the following types of data:

- Data acquired in the course of conducting experiments or examining natural phenomena, or in the course of performing tests according to prescribed procedures;

- Data which describe the characteristics or performance of a natural phenomenon, a material, a device, or a component; and

- Data which instruct, guide, or aid skilled or semi-skilled persons in the proper use, maintenance, or replacement of artifacts, or in techniques and procedures.

The scope and diversity of these activities preclude an explicit listing of inclusions and exclusions of specific data activities; therefore, the following criteria were used to guide the determination as to whether or not a type of data or data activity was within the scope of the census effort:

- Data generated in any of the basic and applied physical, biological, and environmental sciences, basic and applied engineering disciplines and related technologies are included within the scope. Behavioral and human factors data generated in the social sciences are also included; data generated in the other areas of the social sciences are included to the extent that the data are used by scientists or engineers engaged in scientific and technical activities.

- Data embodied in any physical format, from magnetic tape through standard reference manuals and handbooks to programmed or other instructional manuals, are within the scope. Oral and other informal media of data communication are not included except as required to characterize important interfaces between formal and informal data communication media or systems.

- Data in all stages of refinement, from raw measurements through reduced and analyzed data to standard reference data, are within the scope.

- Data in the public domain are within the census scope; in addition, other data are included, if potentially available to the scientific and technical community. Data held by Government intelligence agencies or other highly restricted data are not within the scope. Proprietary data held by private organizations, but made available for external use under appropriate conditions, are included; but private data are excluded.

- Data activities involving either the collection, reduction, analysis, storage, retrieval, analysis, or dissemination of data are included within the scope. Activities predominantly involving the abstracting, subject indexing, or other handling of research reports and other low data content documents are not included.

- Data activities of national scale are included within the census scope. Data activities serving a regional or a specialized mission are within the scope if the use of the data activity is of national importance. Data activities of an international scope are included if they impinge significantly on national level activities. Data activities of only local scope and without existing or potential national importance implications are excluded, except as specimen cases of local scale data activities which, in the aggregate, are of national significance.

- Data activities located within or sponsored by government, the professions, non-profit organizations, or commercial firms are all within the census scope, except data activities devoted exclusively to formal instruction in colleges and universities.

The above scope delineations can be summarized as follows: Scientific and technical data activities which have not been examined previously in any broad-scale systematic manner, but are potentially amenable to coordination for the purpose of improving our national scientific and technical posture.

Structure of Census Effort

Since the resultant product was intended to guide the formulation of national policy, the census effort requirement was broad in scope and of a summary nature. In addition, no precedent existed for the conduct of such a broad-scale census of scientific and data activities. Consequently, the effort, by necessity, involved development of structuring and inventorying concepts for scientific and technical data and data activities. Since this census effort is a pioneering endeavor, it must be expected to be coarse and incomplete with respect to detail. However, it should achieve the important objective of revealing patterns and trends important in the national context. The selected approach achieves this objective; in addition, it provides a structure on which other, more definitive studies and purposive actions can be built.

At the broader level, the concept of "community of interest" proved to be helpful in structuring the census effort. By definition, a community of interest exists when individuals and/or organizations identify with a common scientific and technical mission, goal, or objective. A normal manifestation of a community of scientific or technical interest is the development of an effort to generate and conserve the data required to pursue the common missions or goals of the community. These data efforts and the larger system of which they are a part may be well articulated and formally structured, or they may be hardly discernible and informally structured. In the communities of interest context, scientific and technical data efforts fall into three major categories:

- Efforts primarily associated with basic and applied research missions;

- Efforts primarily associated with design, development, and test missions; and

- Efforts primarily associated with production, operation, maintenance, and training missions.

In the first category, a community of interest evolves in conjunction with a mission to conserve and advance the scientific and technical knowledge in a discipline such as chemistry or a sub-discipline such as analytical chemistry. In the second category, the community of interest associates not only with specific developmental disciplines, such as aeronautical engineering, but also with specialized fields of development such as spacecraft design and developmental or clinical testing of drugs. In the third category, the community of interest is formed along industrial classifications such as transportation and metal fabrication, or around an applied profession such as medical practice.

The community of interest model has particular merit in making visible the motivational patterns that lend meaning to the structure and functions of data activities associated with each scientific and technical mission. This essentially social model also accommodates the important dynamic functions of data conversion and transfer processes. This model displays the structure of communities and enhances the opportunity to identify meaningful patterns related to data activities within the community. Various communities of interest have developed data communication activities which utilize the following channels or, stated in other terms, operate in one or more of the following system modes:*

Generator ↔ User
Generator → Document Publisher → User
Generator → Document Publisher → Document Processor ↔ User
Generator → Document Publisher → Data Processor ↔ User
Generator → Document Publisher → Document Processor ↔ Data Processor ↔ User
Generator ↔ Data Processor ↔ User
Generator ↔ Data Processor → Document Processor ↔ User

*Arrows indicate directions of major flows of data.

Previous studies of the Task Group on National Systems have covered those three communication channels of system modes shown above which do not involve a data processor component. Therefore, the type of data activity most central to the census objective is that which includes a formal data processing component or effort. The visibility of formal data processors such as data collection networks, data storage and retrieval centers, etc. provides identifiable focal points for census efforts. However, it is recognized that such formal data efforts or data processors represent the intersection of two communities of interest. One is concerned with advancing a particular scientific or technical mission and the other concerns the mission of attaining more effective means of handling scientific and technical information and data. A census effort which was limited to coverage of the data efforts and thereby excluded the broader community of interest which they serve would not fully meet the objectives of the census. Therefore, a census approach was selected which provided for assembly of:

(1) Information concerning data activities as conducted within broad communities of interest, such as a discipline or technology;

(2) Information which characterizes specific types of formal data efforts or processing operations; and

(3) Information which characterizes the elements of data activity found in either specific data efforts or in the broader context of the data activities serving a specific scientific or technical community.

The scope and diversity of information enumerated above preclude use of a single means of assembling and presenting the total census. Information in category (1) is not readily amenable to comprehensive, in-depth censusing of a quantitative or analytical nature. Therefore, the census approach chosen was development of descriptive write-ups which show only the gross characteristics of these broad-scale data activities. In contrast with Category 1, information in Category 2 is more amenable to quantitative and analytical treatment. The approach chosen to collect and present this class of information, therefore, follows normal census practices. Within the census budget allocated, relative little effort could be directed specifically to inventorying of

the individual elements of data activities and efforts. It was necessary to limit the direct examination of these elements to a set of census probes. However, an awareness of these elements was incorporated into the approaches used to collect and structure the other categories of census information.

Table i-i outlines the general methodology used to assemble and structure the census information.

TABLE i-1

METHODOLOGY FOR DEVELOPMENT OF CENSUS OF DATA ACTIVITIES

Work Objective	Method of Accomplishment
Identify and acquire literature describing current status of data activities.	Scan announcement bulletins of document centers, search through document storage and retrieval systems, trace citations in key documents.
Identify key organizations and individuals concerned with data activities.	Personal interviews, literature reviews, and workshops with leading data specialists.
Identify current data activities in the various sciences and technologies and in the different phases of these sciences and technologies.	Draft write-ups describing the data characteristics, data flow, data efforts, and issues associated with each area of scientific and technological effort.
Compile census-like facts currently available for formal data efforts.	Extract census information from documents and interviews and record in worksheets.
Verify the accuracy and completeness of descriptions and census facts about data activities and formal data efforts.	Expose preliminary findings in interviews and workshops with leading data management specialists.
Generate comprehensive write-ups covering selected communities of interest within scientific and technical data activity.	Integrate contributions from interviews and workshops into final write-ups of current status of data activities in the various areas of science and technology. Conduct limited surveys to probe selected data activities.
Structure and analyze preliminary census of formal data efforts.	Survey formal data efforts by mail questionnaires and by facility visits. Prepare directories and tabulations of characteristics of formal data efforts. Analyze information assembled and relate to national system requirements.

PART A

CURRENT STATUS OF DATA ACTIVITIES IN SCIENCE AND TECHNOLOGY

I. STRUCTURE AND CONTENT

The subject of analysis in Part A of this volume is the diverse data activities associated with the substantive activities in science and technology. Two sections comprise this part of the volume; Section II giving analytical descriptions of the data activities in ten selected fields of science and technology, and Section III presenting an overview of data activity in basic, developmental, and applied fields of science and technology.

In Section II, the objective in selecting the ten fields of science and technology was to provide an adequate representation of activity in engineering, as well as in the physical, life, and earth sciences. Another goal was the representation of data management in mission-oriented, industry-oriented, and discipline-oriented fields. Table I-1 shows how this was achieved.

To assist in correlative analysis of the ten descriptions of data management in the selected fields of science and technology, a common structure was adopted as the basis for the format. It should be noted that a consistent nomenclature system is not used in all write-ups. Rather, each field is described in terms appropriate to the specific field. In each of the descriptions, the first main heading is an introduction which defines the field and relates the importance of the data used in the field. The second subsection concerns data characteristics. This section classifies and characterizes the data by functional use, discipline, or measured properties. The third subsection of the write-up, pertaining to data flow, is concerned with the users, generators, and intermediaries associated with the communication of data. It cateogorizes the users of data, indicating who the users are and how the data are used. It then categorizes the primary data generators and the data communication intermediaries. The fourth and final subsection summarizes some typical problems relevant to data management in each specific field of science and technology, making suggestions for resolution of certain ones.

A cursory examination of the findings resulting from the survey of ten selected fields of science and technology leads to three primary conclusions concerning the management of scientific and technical data activities:

- Data activities are best understood when viewed in the context of the scientific or technical mission which they serve; consequently, consideration of data management requirements from this perspective is an effective approach.

- Commonality of data characteristics and data flow is defined more by the type of data activity (discipline-research, mission-development, applications-product) than by the field of science or technology.

- While data characteristics (form, volume, quality, rate of obsolescence, value, etc.) and data flow needs and patterns are highly interrelated, separate analyses of these factors are useful for identifying data system requirements. For example, consideration of data characteristics leads to definition of requirements for data management systems; whereas consideration of data flows leads to definition of requirements for data handling systems.

Elaboration of these three findings is the essence of the generalized assertions set forth in Section III. of Part A of this volume. This section (Section III) is concerned with the data and data flow characteristics associated with basic research, developmental, and application phases of science and technology.

As a set, the surveys in this part of the census begin to delineate the commonalities and differences which exist among the different fields and phases of scientific and technological activity. Such examinations appear vital to the establishment of a base of understanding to support the future evolution of new and improved data management and data handling systems. Fortunately, more definitive examinations have already been initiated in a few areas; hopefully, means will be found to continue and expand this vital activity.

TABLE I-1

COVERAGE OF THE WRITE-UPS

Subsection	Field	Primary Orientations
A. Aerospace Science and Technology	Physical Sciences & Engineering	mission & industry
B. Electronics and Electrical Engineering	Physical Sciences & Engineering	industry & discipline
C. Materials Science and Engineering	Physical Sciences & Engineering	industry
D. Chemistry and Chemical Engineering	Physical Sciences & Engineering; Life Sciences	industry & discipline
E. Agriculture and Food Technology	Physical Sciences & Engineering; Life Sciences	industry
F. Biomedical Science	Life Sciences	discipline
G. Pharmacology	Life Sciences	industry & discipline
H. Behavioral and Social Science	Life Sciences	discipline
I. Environmental Science and Geosciences	Earth Sciences	discipline
J. Oceanography	Earth Sciences	discipline & mission

II. CURRENT STATUS OF DATA ACTIVITIES IN TEN SELECTED FIELDS OF SCIENCE AND TECHNOLOGY

A. Aerospace Science And Technology

1. Introduction

Aerospace science and technology involves a multi-disciplinary effort ranging from basic scientific investigation through the research, development, test and evaluation of systems to the operation and maintenance of vehicles. The field embraces virtually every scientific and technical discipline, including chemistry, life sciences, mechanical engineering, and data processing, but the most significant of these is electronics. Electronic components and systems account for nearly half of the value of the aerospace industry's products, and the industry, in turn, consumes two-thirds of the electronics industry's output. The aerospace industry embodies the nation's largest single group of manufacturing employers, employing 1,407,000 persons in 1967 (the bulk of this employment, 54.1% or 761,000 persons, consists of production workers). Scientists and engineers account for 17%, and technicians another 7%.

Total aerospace sales were $27.3 billion in 1967, which represented about 3.8% of the $700 billion gross national product and a 13% increase over previous year sales of $24.2 billion. This output can be divided into four product categories: aircraft, $15.3; missiles, $4.5 billion; space systems, $5.2 billion; and non-aerospace applications of the technology (e.g., oceanographic, desalination, systems analysis, rapid transit, urban problems), $2.35 billion. (Figure II-A-1)

The national defense implications of aerospace activity are obvious. Sales to the Department of Defense (DoD) in calendar year 1967 were $15.9 billion -- $10.4 billion for aircraft, $4.5 billion for missiles and $1 billion for military space programs. About 12% of the current aerospace employment is tied to the Vietnam conflict, and this effort accounts for about $3 billion in helicopters, fighter and attack aircraft.

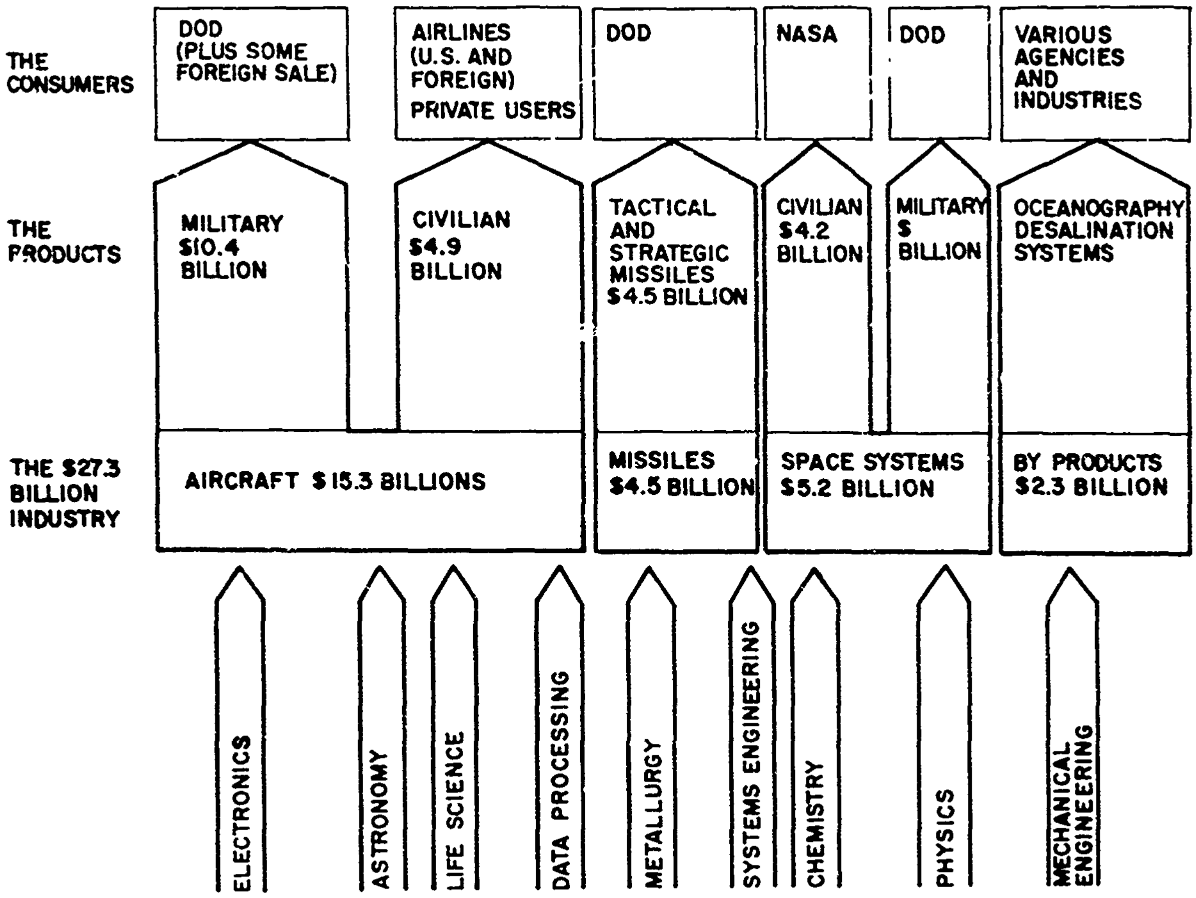


Figure II-A-1 <u>Multidisciplinary Support of Aerospace Effort</u>

2. Characteristics of Aerospace Data

Aerospace data can be divided into two general categories: (1) engineering data that are used solely for design, development and operation of systems and thus form a data flow essentially limited to the industry itself, and (2) data that are "consumed" by users outside the industry. This latter category can be further subdivided into basic science data (astronomy, physics, biosciences, lunar and planetary studies, solar investigations) used to create a coordinated picture of the universe and what the National Aeronautics and Space Administration (NASA) termed "applications" data, such as meteorological geodetic and earth resources data. A third subdivision of "consumable" data, classified military information gathered by secret satellites, is excluded from the scope of this study.

Engineering Data are involved in all aspects of research, development, test, manufacturing, assembly and checkout, logistics and operations. These data constitute the common denominator to the development of all elements of the aerospace system because they form the basic communication link and provide the record of events.

Included in this data spectrum are data contained in. systems analyses and research reports, specifications, engineering drawings and associated drawing lists, inspection and calibration requirements data, equipment logs, technical infomation file data, training and equipment planning documents, configuration control documents. facilities support data specifications, qualitative and quantitative personnel requirements, assembly and checkout and procurement documentation, test support and maintenance materials and operational technical manuals.

Even though system development techniques have evolved in recent years to facilitate simultaneous development of several elements of a weapon system so that total development time may be reduced, the data requirements of a specific system development program usually follow a chronological progression. A performance requirement is first established (e. g. , land a man on the moon in this decade,

redress any imbalance in strategic missile forces vis-a-vis the Soviet Union, develop a supersonic commercial aircraft that will compete with a foreign version, etc.). The steps which follow involve functional system specifications to meet performance requirements, engineering specifications and other types of data. The flow of these engineering development data for typical aerospace systems is described in a later section.

Science and applications data differ from those involved in the engineering function in that they are analogs of physical phenomena and therefore, are in much less refined form. While the division between science and applications has been made for the sake of convenience, it should be remembered that a cloud cover photograph or radiometric map generated by a Nimbus weather satellite is just as much an analog as a stream of electrical impulses from an Explorer satellite describing the flux of solar particles. Each form of data is processed substantially by professionally trained analysts before the data can have any economic or scientific value.

In the case of scientific data, sensors on spacecraft convert physical properties such as temperature, charged particle energy, or magnetic field intensity into electrical quantities. Signal conditioning circuits aboard the spacecraft work directly with these sensors to simplify processing and telemetering of the electrical quantities. Additional processing circuits count pulses, measure the amplitudes of pulses, and measure time intervals to further aid telemetering. Data received by tracking stations are returned to the appropriate NASA center. stored in archives, and made available to the scientific community.

In 1967, NASA's Goddard Space Flight Center, which has responsibility for many scientific satellites, reported it was receiving an average of 237 million data bits per day from such satellites as the Interplanetary Monitoring Platforms, Orbiting Solar Observatory, Orbiting Geophysical Observatory, Orbiting Astronomical Observatory, Applications Technology Satellites and Biological Satellites. This is nearly double the 1966 figure.

For administrative purposes, NASA breaks down its scientific program into lunar and planetary, and astronomy and physics projects, including planetary atmospheres, astronomy, solar physics and biosciences.

Lunar and planetary projects involve study of the condensed material of the solar system. They include earth-based measurements of the electromagnetic radiation from the moon and planets, simulation and terrestrial-counterpart studies, investigations of chemical-mineralogical composition and genesis, and spacecraft observations. It is this latter item that attracts the most attention and generates the most raw data. Some examples of the data are those contained in photos of the moon and Mars taken by Ranger, Surveyor, Lunar Orbiter and Mariner, and the lunar soil constituency data gathered by Surveyor. The volume of these data is expected to increase enormously if the administration and Congress approve further planetary probes, particularly those that would land a capsule to search for life on Mars and measure the soil.

Planetary atmosphere studies, which deal with the atmosphere above 18 miles of the Earth and other planets, have generated data such as temperatures of the isothermal region (above 180 miles), electron temperature in the F-region of the ionosphere, hydrogen and helium constituents in the outer atmosphere, the hydrogen geocorona forming the outer region of the atmosphere, the role of atomic and molecular oxygen and nitrogen in the airglow processes, meteoroid populations--all on earth--and atmospheric pressure and carbon dioxide constituents of the Martian atmosphere and temperature profiles of the Venusian atmosphere.

Particles and field investigations via spacecraft began in earnest with Dr. James van Allen's discovery of Earth radiation belts and has since accelerated to produce data concerning the energetic plasma stream from the sun (generally called the solar wind) and its interaction with the earth's magnetic fields, the various cosmic rays (stellar and galactic) and various other radiation sources in space.

The term ionospheric and radio physics derives from the role the ionosphere plays in reflecting radio waves (as well as shielding the earth's surface from lethal solar radiation). Data collection has concentrated on the previously unmeasured altitudes above the F-region (180 miles), but other work is proceeding on what is called "sporadic E," a thin layer of ionization associated with wind shears in middle latitudes and with the electrojet current over the magnetic equator.

Space astronomy involves the process of using orbiting telescopes and other instruments operating at non-optical wavelengths above the turbulence of the earth's atmosphere to collect analyzable data. NASA divides this program into solar astronomy, and stellar and galactic astronomy.

Solar physics differs from solar astronomy in that the sun is not studied as a star, but for its basic physical properties - much like the study of terrestrial weather. Among the data used are measurements of the ionized iron and calcium atoms in the solar corona, the migration of subphotospheric magnetic currents toward the solar equator over 22-year cycles and sunspots and flares.

Bioscience programs have four data gathering goals: (1) to determine if extraterrestrial life exists anywhere in the solar system and, if so, to study its origin, nature and level of development; (2) to determine the effects of space and planetary environments on earth organisms, including man; (3) to determine the design requirements of life support and protective systems for extended manned space flight; and (4) to develop the basis for fundamental theories in biology relative to the origin, development, and influences of the space environment. Data are gathered using manned space craft and in the Biosatellite series. Further bioscience data gathering is planned for the Pioneer satellite series, which is currently limited to studies of solar particles and fields.

Besides space science data, the other consumable class of data is characterized as applications data. Tiros and Nimbus weather satellites, which are operated under a joint program between NASA and the Commerce Department's Environmental Science Services Administration, are used to collect these data. These satellites have returned cloud cover photos and infrared pictures taken by radiometers. Cloud cover photos take two forms: high-resolution photos taken by the advanced vidicon camera system that can be received and processed only by very sophisticated equipment and the low-resolution APT (automatic picture transmission) photos that can be received by private users.

Of all the types of scientific data considered here, the cloud cover photos provide the only major data used in a "real time" mode. Their value is a function almost entirely of their timeliness, especially in the case of the hurricane season. Real time data are also required in huge quantities for implementation of manned space flight missions, although this use of scientific data applies almost solely to the further development of the manned spacecraft. Exceptions to this rule are the scientific experiments conducted in the Apollo and Gemini missions.

With the broad class of applications data, there is another category produced by satellites, geodetic and navigation data. NASA's Geos program is providing data to refine known distances between any two points on earth to less than 10 meters. The Navy's Transit navigation satellite, which was recently partially declassified and made available to non-military users, provides position data to ships and ultimately will do the same for aircraft.

To describe the characteristics of the data under consideration here, the division between engineering data and science and applications is helpful (Table II-A-1). As has already been noted, the volume becomes progressively greater for engineering data as the system moves from concept to hardware. This volume becomes cumulative as the test reports and specifications for each component accompany the subsystem and on to the final system. By the time a launch vehicle reaches Cape Kennedy or a missile is installed in its silo, a great body of test data has been accumulated.

The process is just the reverse for scientific and applications data. Large numbers of cloud cover photos are analyzed to get the answer to the question, "Will it rain tomorrow?" Millions of data bits are accumulated on particle fluxes to construct a model of the earth's magnetosphere. The reduction in volume in these cases is essential to the understanding and practical use of the data.

Both classes of data have the same relative degree of refinement and technical sophistication. As the data are refined from a scientific space mission or as operational requirements are translated into subsystem specifications, the technical sophistication increases accordingly. If the data flow is conceived as a "bottom up" process, then the refinement at any level ideally matches the requirement of the user--circuit designer, test engineer, contract administrator, system integration manager, university researcher, weather forecaster or NASA administrator.

Orientation also differs markedly between the engineering and science/applications data. The former is almost entirely mission-oriented; data are generated for the sole purpose of supporting a specific system development program, and there is little consideration at the data-generating level regarding possible application in other programs. Conversely, science

TABLE II-A-1

DATA CHARACTERISTICS CHART

	Type of Data	Volume	Degree of Refinement and Technical Sophistication	Orientation	Value	Useful Lifetime
Engineering Data	Performance requirements	Low	Very Low	National goals	Prerequisite low competitive position	Varies with magnitude of requirement
	Functional specifications	Moderate	Moderate	Moderately mission-oriented	Essential to project	Short
	Hardware specifications	High	High	Highly mission-oriented	Valuable for proprietaries' rights	Varies
Science and Application Data	Science	Initially high in data flow	Initially low	Universal	Varies, but usually limited to science community	Of timeless value to historical analysis, but usually superseded by new data.
	Application	Initially high in data flow	Initially low	Practical needs	Great potential	Very short

and applications data are potentially universal. They describe phenomena of wide interest and are disseminated to anyone desiring them. This has been a particular strength of the civilian space program, inasmuch as it has been able to attract international interest through free data exchange.

The orientation of data greatly influences its economic value. The specificity of the engineering data generated in development programs is the cause of its economic value to the company working on the program. Competitive position is maintained by retaining as much data as possible. Universality of potential use influences the value of science/applications data; value increases proportionately with the number of users. The more scientists who can get the basic data, the more scientific analysis can be performed, and the lower the cost of the additional copies of the data needed by each.

Timeliness is clearly a factor in weather and navigation data. It is also true to a lesser extent in engineering data. With the rapid strides being made in technology, it is essential that any developing organization--governmental or industrial--keep abreast of new developments. Data of electronic techniques, for example, obsolesce very quickly (i.e., requirements move from discrete solid-state components to integrated circuits). The key point here is that competitive position is maintained within the aerospace industry through anticipation of future requirements and accumulation of sufficient data on new technology before the need arises.

3. Aerospace Data Flow

Data flow in the aerospace field follows patterns which are different for engineering and science/applications data. As mentioned earlier, engineering data flow is generally managed to meet the requirements of specific system development programs, whereas science/applications data flow is less regulated and is oriented toward the less precisely defined needs of the user communities involved.

Engineering data flow in the aerospace industry matches the requirements of the research, development, test, and engineering cycle (See Figure II-A-2). The first data requirement is a performance requirement, as mentioned earlier.

The next logical step is establishment of functional specifications to meet the overall requirements (develop a launch vehicle or vehicles and spacecraft capable of the lunar trip, build missiles of varying ranges and payload capabilities that will be available in the required time periods, select an optimum speed, size and passenger-carrying configuration for a competitive SST, etc.). At this point, procurement and other support data enter the data stream.

In the final major step in this sequence from the general to the particular, hardware specifications are issued (proceed with the construction of the Saturn V and Apollo, develop a missile force consisting initially of Atlas and Titan to be phased out later in favor of Minuteman and Polaris and Posidon, approve the Boeing airframe and General Electric engine for the SST, etc.). Data at this stage are generally embodied in requests for proposals (RFP's) issued by the cognizant Federal agencies and then established in the basic contracts and later modifications that will determine the relationships between the agencies and industrial firms.

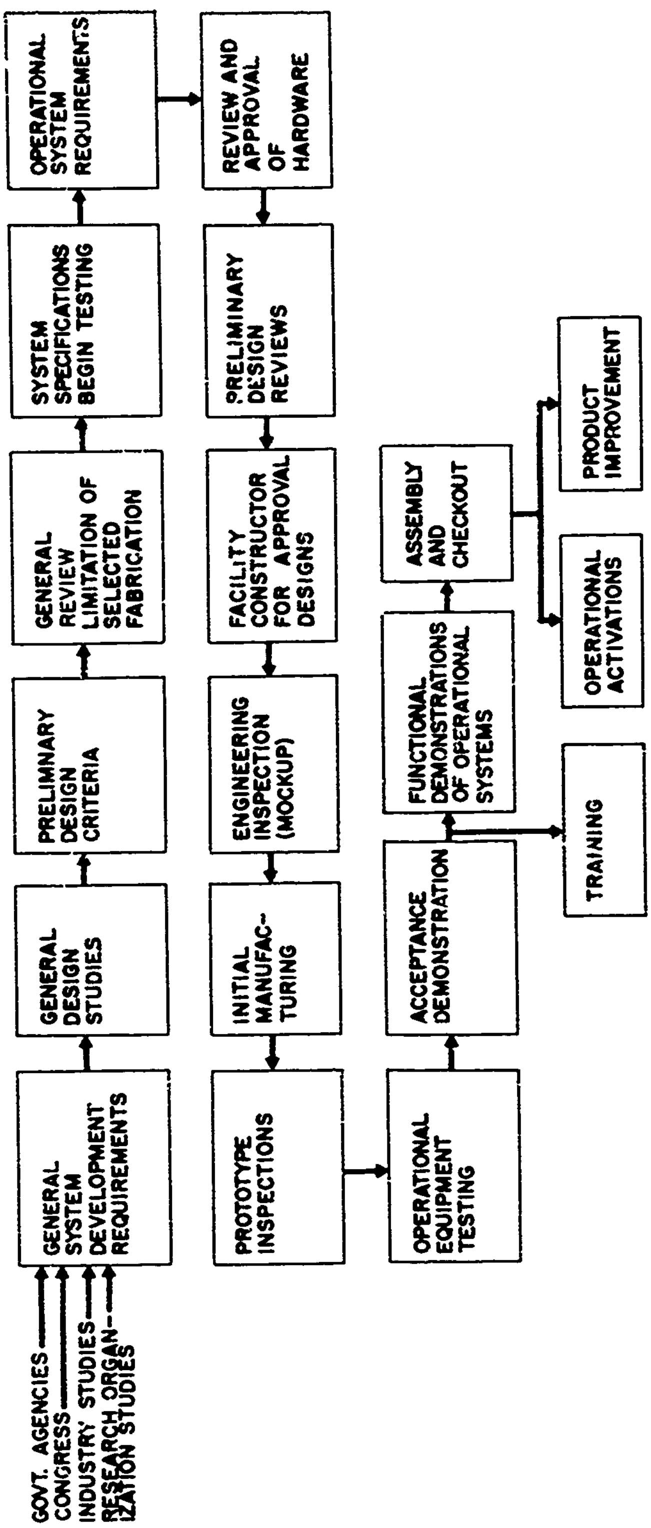


Figure II-A-2 Data Flow Pattern in Aerospace Systems Development

These are not three discrete steps, however, and should be viewed more as a continuous flow, as shown in Figure II-A-2. Furthermore, under the phased program planning initiated by the Department of Defense and adapted by other government agencies, many programs reach only the initial phases before they are scrapped as unfeasible.

To understand the engineering data requirements involved in the development of an aerospace system, it is useful to further break down the three major chronological steps into subroutines that each have unique data requirements. For this purpose, the following 19 steps have been chosen: (1) Establishment of general system development requirements; (2) Beginning of general design studies; (3) Preparation of preliminary design criteria for basic testing; (4) Establishment of general plans and initiation of selected equipment fabrication; (5) Development of system specifications and beginning of testing; (6) Identification of operational system requirements; (7) Review of approval of hardware and facilities recommendations; (8) Preliminary design reviews; (9) Facility construction initiation on approved designs; (10) System development engineering inspection (system mock-up); (11) Initial manufacturing on approved designs; (12) Prototype inspections; (13) Operational equipment testing; (14) Acceptance demonstration of first article; (15) Beginning of training; (16) Functional demonstrations of operational systems; (17) Assembly and checkout, and weapon system acceptance demonstrations; (18) Operational activation; and (19) Product improvement.

As the system evolves from concept to hardware, the amount of data needed at each step increases accordingly. The first step, for example, consists only of a requirements document and this is abstractly worded to avoid inhibiting inventiveness. As the program moves toward general design studies, the data output becomes a series of reports and recommendations. Preliminary design criteria involve R&D procedural data requirements and human engineering criteria for analysis and planning.

Once the general plan has been established, a data explosion occurs involving operational plans, maintenance plans, logistics plans, detail studies and R&D drawings. This step flows naturally to the development of a system specification tree including performance specifications and reliability data, and procedures to support preliminary design and R&D testing. Operational system requirements include functional performance specifications, operational functional analyses, support functional analyses and maintenance analyses for the system and initial hardware, facility, personnel and support data recommendations.

The hardware and facilities recommendations are then reviewed and the following data are generated; hardware identification sheets containing design requirements and recommended solutions, personnel requirements,data in preparation, and operational procedures and drawing identification planning.

At the preliminary design review stage, a vast amount of data is generated on requirements for hardware, facilities, personnel and technical data. A few examples are functional flow diagrams, preliminary design criteria, reliability calculations, cost effectiveness data, study reports, site activation drawings, proposed facility drawings, time-line drawings of job operations, qualitative and quantitative personnel needs, training equipment planning information, proficiency evaluation and training plan, approved equipment lists, assembly and checkout plans, test plans, and technical data requirements index.

When facility construction is begun on approved designs, design drawings and planning documents are required. At the mockup stage, proposed technical manuals are composed to cover operational and maintenance procedures, trainer performance specifications are drawn up along with

training course outlines, and safety criteria and test plans are formulated. As the program enters the manufacturing phase, data needs include manufacturing plans, engineering drawings and associated data lists, production planning sheets, acceptance test procedures, manufacturing and quality control records, and special tests and test equipment drawings and procedures.

At the time of the prototype inspection, data include engineering drawings, preliminary model specifications, preliminary procedures, spares provisioning data, acceptance test procedures, preliminary model specifications, test directions, and preliminary trainer model specifications.

As operational equipment testing begins, technical data are generated on assembly and checkout procedures, detail test directives, functional test procedures, reliability, configuration control, operating and maintenance procedures, engineering drawings, logs, instrumentation data, and data evaluation sheets. Essentially, the same data are required at the time of acceptance of the first article.

Training, which does not have to adhere rigidly to this chronological order and can begin almost any time, involves training courses, preliminary technical manuals, training aids, manning documents, equipment operating and maintenance procedures, and proficiency training and evaluation instructions.

Operational system functional demonstrations involve integration of acceptance test procedures, technical manuals, operational and maintenance checklists, inspection and maintenance checklists, inspection and maintenance work cards and sequence charts, engineering change proposals, operational readiness training courses, personnel subsystem testing plans, logs, and failure data.

Assembly and checkout require those assembly instructions not included in the technical manuals, acceptance functional test procedures, handling and transporting instructions for peculiar checkout equipment, maintenance support data for this equipment, the technical manuals, data record sheets, work cards for check-out and control sign off, assembly and checkout sequence cards, and acceptance demonstration criteria.

In the final two stages, operational activation and product improvement data consist of technical manuals, operational and maintenance checklists, inspection and maintenance work cards and sequence charts, engineering change proposals, logs, failure data, proficiency evaluation and unsatisfactory reports.

Another view of engineering data flow, which provides perspective concerning the relationships between research and development data, is based on examination of four flow modes characteristic of the aerospace field. These are planning data flow, research data flow, developmental data flow and production data flow. While these four modes comprise a chronological sequence involved in an overall aerospace systems project, each represents a discrete mode.

Planning data flow begins with the mission-oriented forecast for R&D on a new product, system or advance in current technology. These data effect control of the selection of the areas to be exploited and thus determine the data required. Flow revolves around narrowly defined boundaries, as shown in Figure II-A-3.

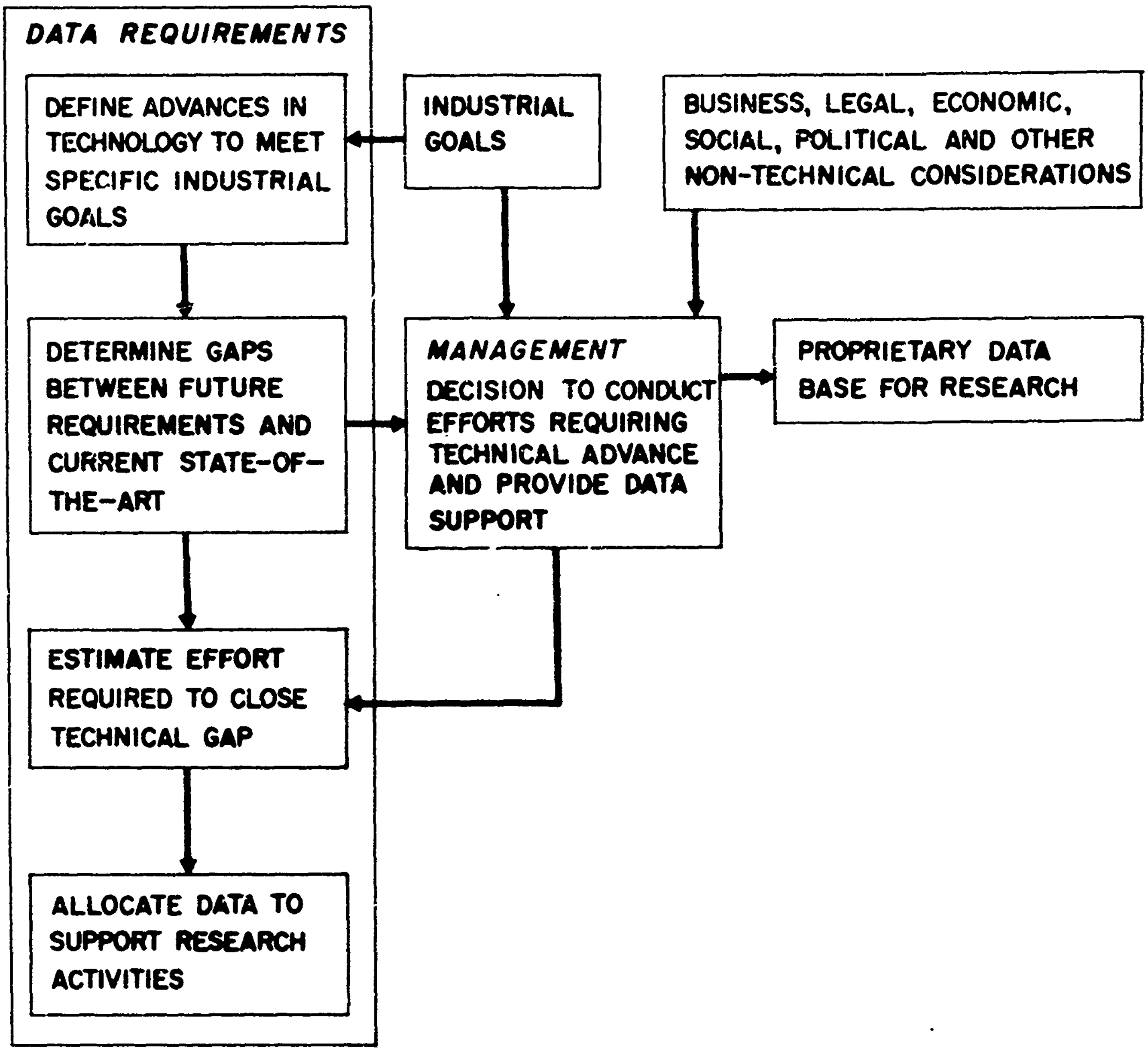


Figure II-A-3 Planning Data Flow

Once the research has been committed, the data cycle still confined within the boundaries of research becomes better defined. The project team gathers pertinent data available and applies efforts toward extending the technology. When all available data resources have been exhausted, the researchers generate new data as the result of their work. These data are evaluated and then published for review. Meanwhile, experimentation generates redundant data that are published internally for validation. The final step is formal publication of results, a process that is estimated to take an average of two to three years after final validation of the research data.(See Figure II-A-4.)

Mission influences begin to have major impact on the program data in the developmental phase. Engineering drawings evolve from the program to determine the final configuration of the system. Internal reports are generated during this activity to describe the operation and authenticate the system for historical purposes. At this stage, the time element plays a major role in regulating the data flow. In a crash program, for example, the data cycle is compressed and a number of intermediary control steps are eliminated. In an extended development cycle, extra steps may be added that slow the data flow but enhance validity. (See Figure II-A-5.)

Production data are similar to those generated in the development phase. These data are generated to produce the prototype and are continuously updated to assure compatibility with operational requirements. Data regarding reliability and maintainability are generated for the user organization, where they are distributed for operational use such as field maintenance and logistics. (See Figure II-A-6.)

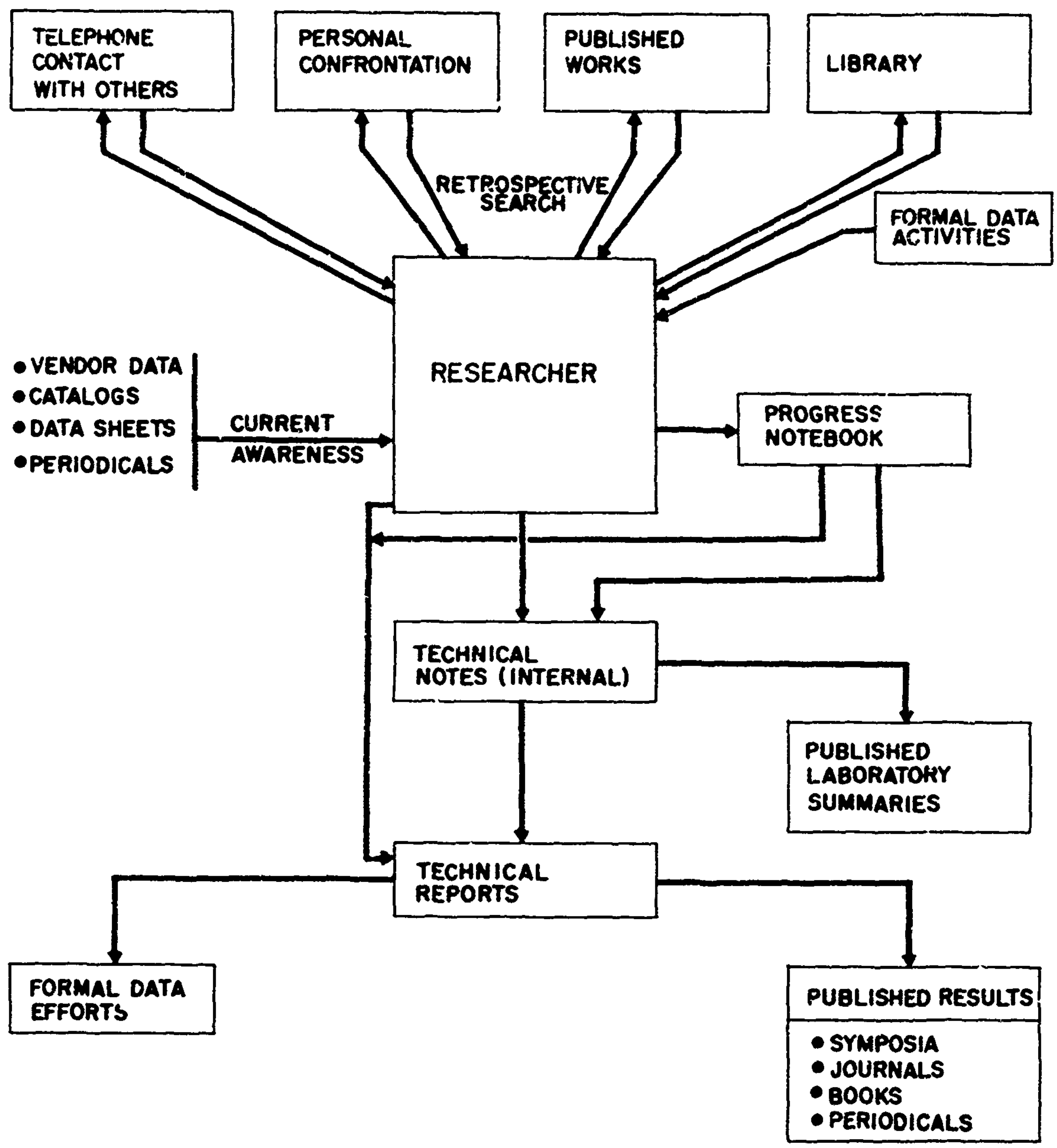


Figure II-A-4 <u>Research Data Flow</u>

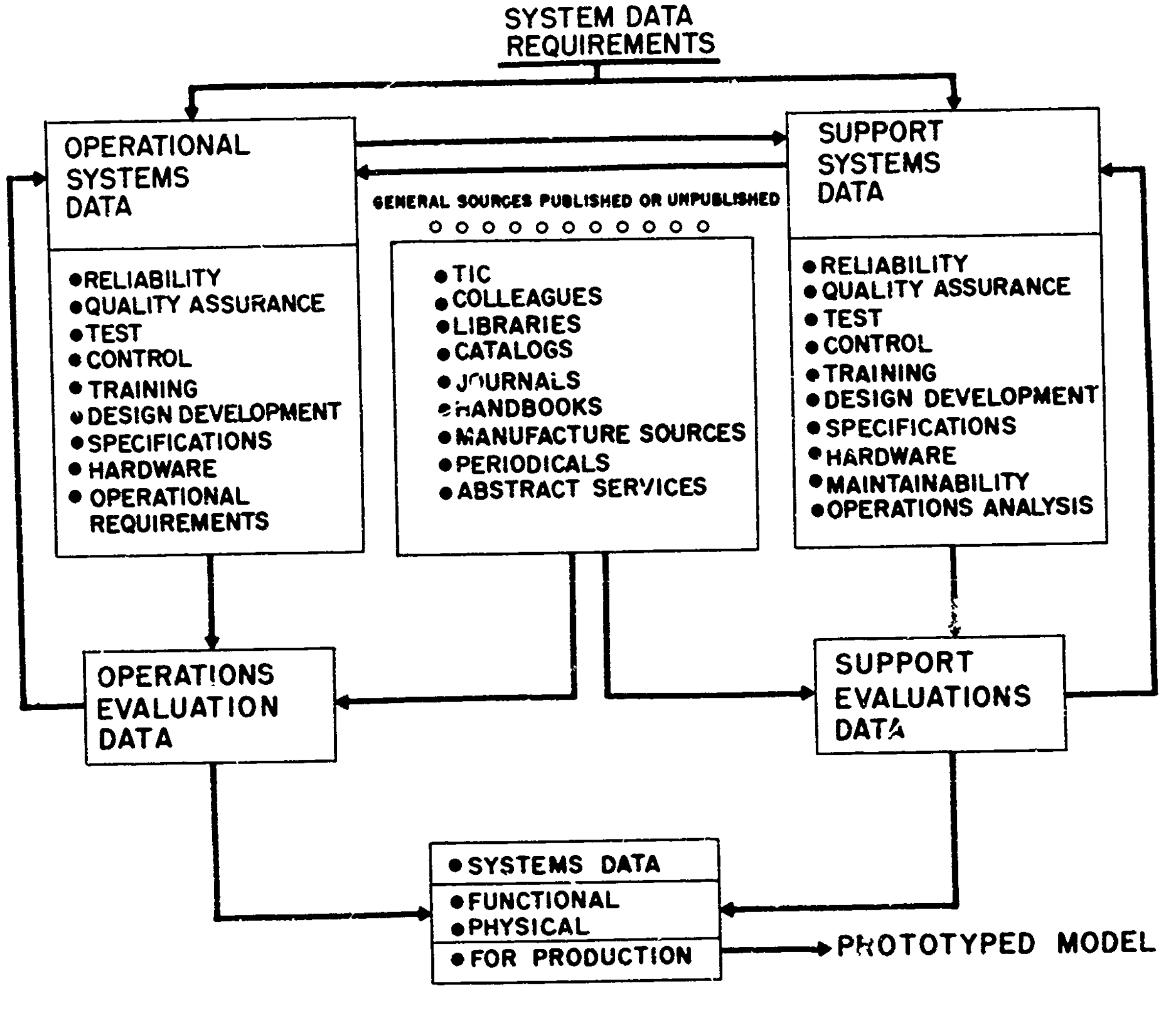


Figure II-A-5 Developmental Data Flow

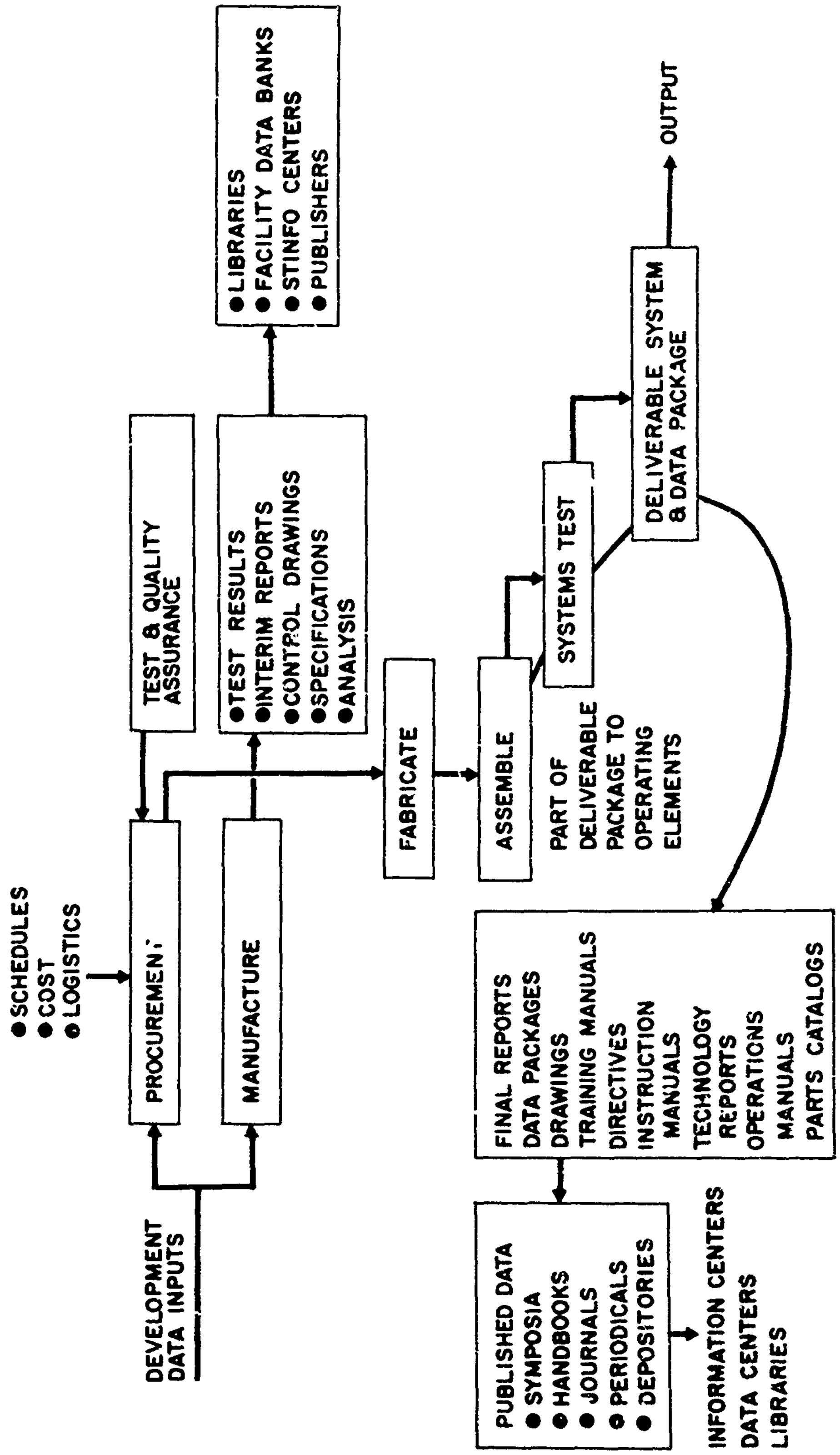


Figure II-A-6 Production Data Flow

It is obvious that there is a relationship between certain types of science/applications data flow and engineering data flow. Data which describe the environmental conditions which constrain or otherwise affect design of aerospace systems enter the engineering data stream. For example, data obtained by the Ranger, Lunar Orbiter and Surveyor satellites concerning the lunar environment are said to have influenced the design of the Apollo Lunar Entry Module.

At the same time, much data from these three and many other sensor platforms (i.e., satellites, sounding rockets, ground based observatories, and balloons) flow to the space science community. This flow, illustrated in Figure II-A-7 for a portion of the NASA program, involves a three-step process. Data are first telemetered from satellites to tracking stations, such as the Deep Space Network (DSN) at Jet Propulsion Laboratory and Satellite Tracking and Data Acquisition Network (STADAN) at Goddard Space Flight Center. The data are transmitted to NASA field centers; Jet Propulsion Laboratory (JPL), Langley Research Center (LRC), Ames Research Center (ARC), and Goddard Space Flight Center (GSFC); where scientific program managers reduce or refine the data. Reduced data are sent to the National Space Science Data Center (NSSDC), where they are stored for future use by the scientific community at the various NASA field centers and other research establishments. Reduced data ultimately find their way into the open literature through technical meetings, reports, and journals. Reports are generally prepared under Federal study contracts, and are therefore available through the Clearinghouse for Scientific and Technical Data and/or the Defense Documentation Center.

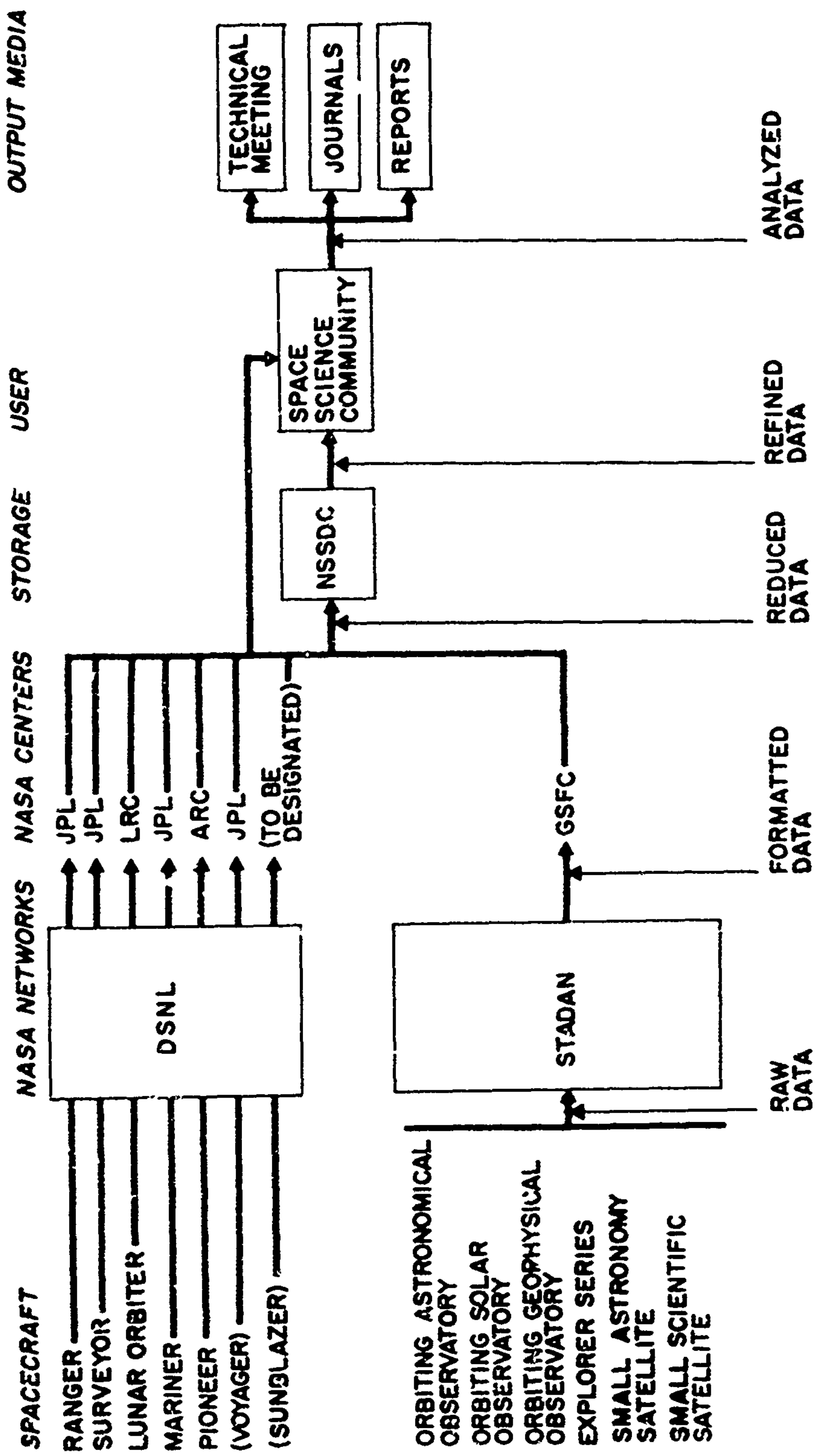


Figure II-A-7 Space Science Data Flow

Data users in the aerospace field can be divided into three broad categories, each with specific, although sometimes overlapping, requirements: government, industry, and the university community. These three broad categories also hold true for data generators, or primary sources of aerospace data, inasmuch as any given organization can be (and usually is) simultaneously a user and a generator. A university conducting a scientific experiment on board one spacecraft generates data to be used in arriving at a better understanding of the universe or aiding in the design of a manned spacecraft; at the same time, the university-based user may consume engineering data generated by earlier programs to aid in the design of other satellite-based experiments.

The data user/generator situation is somewhat more complex in government and industry. A convenient way to categorize these is to divide them into primary and secondary participants in the aerospace data flow. In the case of the government, it is relatively simple to identify the primary user/generator organizations; these are the Defense Department, particularly the Air Force, which has the responsibility of using aerospace data to satisfy its national defense mission; NASA, which was chartered in 1958 to advance aeronautical and space technology for peaceful purposes; and the Federal Aviation Administration, which sponsors the major research and development function in support of operational commercial aircraft. Each of these three agencies requires specific types of engineering and science/applications data. In many cases, these needs overlap. NASA's Apollo and the Air Force's Manned Orbiting Laboratory (MOL) program, for example, use almost identical types of data relative to life support systems, re-entry, electronics reliability, propulsion, materials technology, etc. The Air Force, NASA, and FAA all have major stakes in supersonic flight.

Outside this relatively homogeneous government aerospace community, there are other agencies that use aerospace data sporadically, although they rarely generate any. Specific examples are the Department of Commerce, which is a partner with NASA in the weather satellite program; the Department of Interior, which participates in the geodetic survey of the moon, and which is planning an earth resources satellite program of its own; the Department of Housing and Urban Development, which has expressed an interest in applying the systems methods and using the advanced technology generated in aerospace programs; and the Atomic Energy Commission, which participates with NASA in the nuclear rocket and nuclear space power generation development efforts and which can almost be considered a primary participant in the aerospace data flow process. These organizations are considered secondary participants because aerospace science and technology are not their primary functions.

It is beyond scope of this study to identify the discrete user groups within these large agencies, but it is important to recognize that they do exist. The NASA headquarters staff, for example, is not a direct user of engineering or science/applications data; its primary function is program management and fiscal control. Again using the "bottom up" data model, the users can be identified at the level at which the data are needed. Engineering data on large liquid rocket engines, for example, are both consumed and generated at the engine and rocket stage project offices at the Marshall Space Flight Center. Because of its special interest in X-ray astronomy, the Naval Research Laboratory is a major factor in the data flow in that field. Among the secondary participants, the Agricultural Research Service has become an important factor in automatic data processing of earth resource photos gathered by aircraft and spacecraft.

A similar situation prevails among industry. Primary firms can be identified as aerospace manufacturers and airlines. Each consumes data (i.e., engineering specifications) and generates data (i.e., requirements) both in their relationship with the cognizant government agencies and among themselves in the contractor-subcontractor and vendor-buyer relationships. Inter-company data flow also occurs when firms form teams to bid for contracts or are jointly involved in projects on an associate prime contractor status. Industrial participants in the aerospace data flow are not limited to profit-making firms; the non-profit organizations such as Rand Corporation, Institute for Defense Analyses, and Mitre Corporation play a key role. These organizations, originally set up to advise government organizations of impending requirements, have grown in stature and presently act as a functional intermediary between Federal agencies and industry. In this role, they both generate data on requirements and evaluate progress reports of hardware producers.

Secondary industry participants in aerospace data flow include those companies and industries outside the field of aerospace science and technology that are beginning to find use for aerospace-generated data. They are almost solely consumers, rather than consumer/generators. Examples abound in NASA's justification of its technology utilization program regarding the impact of these data on other segments of industry. Data which are associated with bearings, welding, quality control procedures, data processing, electronic components, even better bathtub caulking compounds and brasieres, have been claimed as part of the "fallout" or "spin-off" of space technology. While the attendant publicity has been viewed skeptically, the impact itself may be expected to grow. To emphasize the impact of this data flow, it is pertinent to note that the aerospace industry itself estimates that more than $2 billion a year (nearly 10% of the total technology output) goes into non-aerospace uses. Specific fields which are increasingly using aerospace data include oceanography, chemical research, medical research, and agriculture This factor is evident in analysis of data flow included in sections of this report to follow.

Science Communication
Washington, D. C. 200 07
COSATI Data Activities Study
Final Report - F44620-67-C-0022

30 April 1968

Intermediaries in the flow of aerospace data can be described as analogous to those in a generalized information system. There is an input function performed by collection networks (e.g., the Deep Space Network, Satellite Tracking and Data Acquisition Network, Space Detection and Tracking System), there is a storage function performed by data centers (the National Space Science Data Center, FAA Aeronautical Center) and data document depositories (Aeronautical Chart & Information Center, Aeronautical Standards Group), and there is a dissemination function consisting of published works (handbooks, lists, journals, reports, compilations, professional meeting proceedings, and technical notes) and informal sources (technical meetings and personal files). While this model applies in most cases as a one-way data flow, the feedback mechanisms in technical meetings and professional journals should not be ignored. Neither should other modes of information communication, which constitute one of the major data channels, even though they are at best difficult to measure.

Examples of each of the three functions are described in the following paragraphs which represent primary elements of data flow in the aerospace field, since it is impossible within the scope of this study to catalog them all.

Three examples of input element. specifically collection networks, are NASA's Deep Space Network (DSN) and Satellite Tracking and Data Acquisition Network (STADAN) and the Defense Department's Space Detection and Tracking System (SPADATS). Each fulfills a portion of the given agency's data-gathering mission - monitoring deep space probes in the case of DSN, collecting data from earth-orbiting satellites in the case of STADAN and cataloging space objects with an eye to national security aspects in the case of SPADATS.

DSN is a contractor-operated facility of NASA run by the Jet Propulsion Laboratory of the California Institute of Technology. It maintains two-way communications with unmanned spacecraft over the distance range from the Earth of 10,000 to several million miles. STADAN performs similar functions for earth-orbiting satellites. It evolved out of the Ministrack network and is operated by Goddard Space Flight Center. Both science and applications data are handled, and output goes to experimenters, operational users (such as weather forecasters), and into the data center, which is also located at Goddard. SPADATS, which is operated by the Air Force with headquarters in Colorado, catalogs all man-made objects in space and reports on their number, size, paths, and life cycles. Volume exceed 7,000 observations made, processed and categorized each day. Output takes the form of an up-to-date catalog of objects in space.

Two examples of storage elements are the FAA Aeronautical Center in Oklahoma City and the National Space Science Data Center (NSSDC) at Goddard Space Flight Center, which are geared to specific user needs. The FAA facility covers accident statistics, aircraft registration, airman certification data, maintenance data and airway charts and maps The NSSDC acquires primarily reduced satellite data, as well as sounding rocket, high-altitude balloon, and ground-based observational data. NSSDC is unique, in that its output can be in either hard copy printed form or in computer-compatible magnetic or paper tape. To facilitate data exchange, the Center publishes a semiannual catalog of experiments (organized by scientific discipline, space vehicle and experiment), a semi-annual catalog of correlative data of comparable ground-based scientific data, data users' notes describing reduced data available from the Center, and various other announcements and bibliographies. The Aeronautical Standards Group maintains data applicable to aircraft design. Some representative titles of handbooks available include "Ground Loads," "Strength of Metal Aircraft Elements," "Aircraft Propeller Handbooks," "Vibrations and Flutter Prevention Handbook," and "Plastics for Flight Vehicles."

Another example of the data archives available in the aerospace field is the vast number of data banks established for the Apollo program. These were developed solely for that program and are not currently available for other aerospace applications. Their inclusion here is to indicate the types of data banks associated with a major development program. Examples, all at NASA's Manned Spacecraft Center in Houston, are Apollo Central Metric Data File, Apollo Drawing Data Bank, Apollo Engineering Microfilm Library, Apollo Failure Data System, and Apollo Test and Reliability Information Center.

Obviously, all the storage facilities listed above also serve as dissemination points. The output of the NSSDC has already been mentioned. In addition to these storage centers, where data can be retrieved on demand, the aerospace field generates published material sent routinely to persons working in the field. The formal NASA scientific and technical information program, for example, generates the following documents, which are rich in data:

- Technical reports containing scientific and technical information considered important, complete and lasting contributions to existing knowledge;

- Technical notes containing information less broad in scope but still considered important as a contribution to existing knowledge;

- Technical memorandums containing information receiving limited distribution due to its preliminary nature, security classification, or other reasons;

- Contractor reports containing technical information generated in connection with a NASA contract or grant and released under NASA auspices;

- Technical translations containing information published in a foreign language considered to merit NASA distribution in English;

- Special publications containing conference proceedings, monographs, data compilations, handbooks, sourcebooks, and bibliographies; and

- the Technical Utilization series Tech Briefs, Technology Utilization Reports and Notes, Technology Surveys and other descriptions of in-house or funded work slanted toward potential industrial users.

Similar programs for routine dissemination exist at the FAA, the Air Force Office of Aerospace Research, the School of Aerospace Medicine and other agencies. Companies also disseminate technical data to a list of interested parties, but these are normally geared to promoting the sale of their own products. They may be regarded as formal data activity, however, since they are organized and influence design of aerospace systems.

Informal sources, although lacking the structured format of formal channels, are often considered of great value within the aerospace community because of their timeliness. This is particularly true of the staff-written news magazines such as

Aerospace Technology and Aviation Week, which short circuit the usual lengthy approval process to disseminate data to potential users while it is still timely. Other, more specialized, publications through which data are presented by researcher-authors provide superior structure and validity at the expense of somewhat decreased timeliness.

Other data sources of major importance are the technical conferences sponsored by the industry itself. Most of these are open and seldom are classified data presented. Sponsoring organizations represent broad segments of the industry, but the major ones are the American Institute of Aeronautics and Astronautics (AIAA) and the American Astronautical Society. As an indication of their broad technical scope, the AIAA in 1967 sponsored meetings presenting data and information on aerospace sciences; flight tests, simulation and support; sounding rocket vehicle technology; structures, structural dynamics and materials; thermophysics; telemetering; marine vehicles; solid propulsion; reliability and maintainability; commercial aircraft; energy conversion; guidance and control and flight dynamics; electric propulsion and plasmadynamics; and missile systems.

4. Aerospace Data Management Problems

The fundamental data requirement in the aerospace field is claimed to be an industry-wide system capable of generating, on demand, up-to-date, validated data with the proper degree of refinement for the user. The premise for this claim is the apparent need for enhanced data flow between the many engineering development and space science/applications programs housed in a multiplicity of organizations. In a utopian system, these data would be created and archived without any inconvenient extra effort on the part of the data generator and retrieved by potential users who have no particular sophistication in the use

of data systems. Such a utopian system, of course, would be fantastically expensive and perhaps could not even pay for itself on the basis of eliminating duplicating efforts. It may, therefore, seem that the present haphazard data channels, such as personal conversations, technical meetings, and industry publications, show the best cost/benefit ratio.

However, based on the assumption that improvement is necessary, it may be useful to describe an ultimate system if only to identify smaller steps that could be taken in the immediate future to provide moderate improvements to data flow at minimum costs. The first requirement for development of such a system is that the data generators, particularly the mission-oriented engineers working under the pressing schedule and cost constraints of given aerospace projects, need not be concerned with formatting the data. It would be unrealistic to expect them to take time out from their principal work,or for companies to allot portions of their contracts to data formulating. Thus, any industry-wide data system would have to be able to accept extremely raw data, or provisions would have to be made for minimal preliminary processing at the source by data processing specialists. This facet of ideal system operation is not as unrealistic as it might appear. Just as commercial bank customers with no knowledge of data processing receive better service through the use of magnetic ink-imprinted checks, use of similar encoding techniques might be beneficial for the aerospace industry to develop. Particularly engineering data thus encoded could be periodically transmitted to a main data center or to regional satellite centers. Transmission could be implemented via mail for drawings and other analog data, and by telephone line for digital data in which timeliness is critical to economic utility.

The most costly and critical element in such a system would be the indexing required prior to input. Unless data can be indexed so that even the most unsophisticated user can find what he wants, the system won't be used and the money will be spent in vain. Specifically, an engineer designing a telemetry system should be

able to query the storage system for all data on a subject as broad as RF amplifiers, or an engineer working on a liquid rocket engine should be able to get test data generated under government contract on properties of nozzle materials, using their respective natural languages for search.

The various test and engineering data would thus be available either from a central depository or from the source itself. This option would appear to be necessary to reduce the time lag in getting data for crash programs. An alternative would be to develop a system of priorities based on the urgency of the project and the stature of the data user. For example, if an anti-ballistic missile defense system were to be found essential to national defense, it is obvious that reduction in delay in the flow of data required for those engaged in accomplishing design would be highly desirable. .

An embryonic version of such an ideal system is embodied in the Interagency Data Exchange Program (Figure II-A-8). This program was established in 1959 by the three military services to prevent duplication of testing efforts in what was then the extremely critical ballistic missile development program. NASA has since joined IDEP. The basic purpose of IDEP is to provide automatic exchange of test data generated in the development of aerospace systems. These data include specifications, summaries of tests scheduled or in process, failure analysis reports (those reports deserving special attention marked with red flags), and general technical reports and papers on the application, reliability, quality assurance and testing. Topics include electronic, electrical, mechanical, and electro-mechanical parts; materials; production processes; pyrotechnic test equipment; procedures; and reliability information.

The program is not large, in relation to the giant aerospace industry. The IDEP office estimates that its 20,000 microfilmed reports cover 30,000 separate items and that another 250-300 new reports are added each month. Cost of accumulating these data is estimated at $50 million.

While IDEP falls far short of the industry-wide ideal system, it does show promise for meeting the requirements for ease of input and output. Participants are not required to generate reports specifically for IDEP, but all component test reports created to fulfill a government contract requirement may be considered suitable for inclusion. The only additional requirement is the preparation of a standardized summary sheet used by many participants for their own internal requirements. Classified and proprietary data are excluded from the system.

At the output end, data are available either from a quarterly report listing arranged by a nine-digit part identification code, or what is called a visual coincidence report indexing system, which consists of a set of perforated cards indexing each report by part type and test environment. In either case, the indexing system refers the engineer to the appropriate microfilm cartridges. Using a microfilm reader-printer, he can locate and scan the report and, if desired, obtain a hard copy of any page. There is no cost to the participant, but a company cannot charge off the time of its personnel using the system to a government contract.

IDEP officials claim that the system has reduced the estimated 20-30% of an engineer's time typically spent on data search and saved the government more than $5 million by not duplicating tests already documented.

The Space Science/Application Flood. In another realm of aerospace activity, as noted previously in the section on data characteristics, the principal problem in managing space science/applications data is one of volume. The vast amount of data returned from scientific satellites requires elaborate reduction and analysis before any clear picture of physical phenomena emerges. This problem is particularly critical with respect to applications data, such as weather conditions, where there is demand for real-time, refined data.

One of the most encouraging aspects of this field has been NASA's efforts to improve on-board data processing. One method is known as previous element coding (PEC), a technique developed for interplanetary spacecraft that uses computer capabilities to sense when a particular piece of data is the same as the one that preceded it. When this happens, the data point is not sent and the ground receiver's logic merely repeats the previous data point. The technique is expected to be particularly useful for Mariner-type photographic missions in which the severe spacecraft power limitations restrict the amount of data that can be returned.

On-board computers are expected to have a major impact in reducing the amount of data telemetered from satellites. Advanced integrated circuits now under development show the promise of hooking up a small, low power computer to each satellite experiment. These computers would return to earth only the significant data.

Another effort aimed at reducing the data deluge overwhelming NASA's scientific satellite program is the growing sophistication of spacecraft designers. Earlier satellites required vast amounts of "housekeeping" data: i.e., information on thermal conditions, power supplies, stabilization, status of electronic components, etc. As more experience with satellites in the space environment is gained, it becomes less imperative to measure the performance of equipment other than the experiments themselves. This does not hold true for radically new spacecraft, but there is a tendency toward fewer of these.

Despite these factors, the amount of scientific data gathered by satellites is increasing as the larger, observatory-class satellites begin to take over. The National Space Science Data Center is playing a major role in handling this data flow and can be expected to grow in importance in direct proportion to future space efforts.

The question underlying all considerations of improving the data flow in the aerospace industry is not one of technology, but cost. The present technology is obviously meeting data requirements, despite the problems. Any steps proposed to improve the present data management will have to have demonstrable cost saving potential.

In the long haul, it seems likely that a concept such as that of the National Space Science Data Center will evolve into the principal data dissemination medium for space science and applications data, and something like the Interagency Data Exchange Program will evolve into the aerospace industry's principal engineering data channel. These organizational entities provide a structural foundation for future systems. Barring a top-level government decision to establish a separate, well-funded data activity, the outlook is for the current institutions to assume greater responsibilities--always lagging somewhat behind needs, but never to the point of crisis.

B. Electronics and Electrical Engineering

1. Introduction

Electronics and electrical engineering, for the purpose of this study, is defined as the field of engineering devoted to the practical application of electro-technology -- the generation or use of electronics and their direct effects or responses -- to useful work. In the past, the field has been divided traditionally into two functional disciplines, electronics and electrical engineering.

The former has been concerned with both electrical power and the generation, transmission and modulation of electrical or electromagnetic signals. Typically, the electronics engineer deals with the problems and equipment employed in sensing and measuring, communicating, storing, or processing data. These functions may be applied for the handling or transmission of information between people or may be employed by people for the direction and control of machines and processes.

Electrical engineering has been considered more narrow in scope, since it deals with the generation and use of electricity or magnetism. It involves the conversion of electrical energy either to or from heat, light, mechanical, or chemical energies, or combinations thereof to accomplish useful purposes. These functions may be applied to electrical-power generation, storage, and transmission; environmental control (heating, cooling, humidifying, lighting); and the operation or manipulation of electromechanical devices.

More recently, the division between electronic and electrical engineering has become less distinct -- with the possible exception of the electrical power specialist. Until the past 6-7 years, the high-power engineer lived in a world apart. Others simply never were exposed to the problems associated with multi-megawatt power generation and transmission. Now we find electronics engineers striving to transmit power through space using millimeter-wave radio energy and we find laser researchers dealing daily with multi-megawatt-pulse and near-megawatt continuous-wave transmissions at optical frequencies. Thus, the emergence during the

past two decades of semiconductor or solid-state devices with their small size and broad flexibility almost catalytically has drawn the two fields into one. We find the electronics engineer integrating exotic power and cooling sources with electronic systems while the electrical engineer depends more and more on sophisticated solid-circuits for the control of otherwise prosaic devices.

The importance of electricity and electronics and the associated data activities can be understood by contemplating the size of the technical community involved. Over 5,000 firms, 150,000 scientists, and 250,000 engineers are either directly or indirectly involved in the business of providing electrical or electronic products and services. Millions of specialists, technicians, and semi-skilled workers support this industry which in 1968 may top $24 billion gross, according to the Electronics Industries Association. Furthermore, it is steadily growing at the rate of $2 billion a year and no leveling off has been predicted for the foreseeable future.

Nearly half the total expenditures in this industry are from the U.S. Government ($10.5-12 billion), and the bulk of this portion is provided by the Defense Department. About one-quarter of the market is derived from high-quality products for the communications, computer and data processing, and industrial equipment sub-industries ($6.2-6.5 billion). The other quarter of the market is provided by consumer sales and replacement parts ($5.5-6.0 billion).

It has been said that in no other industry are so much data generated, so much data disseminated, and so much cooperation maintained among industrial, university and government contemporaries. It has also been claimed that, because of the complexity of the technology and the breadth of the industry, in no other field is there so much duplication of effort. The very same data may be generated for use in developing geophysical sensing, data communication, and fire-control systems without any apparent correlation or communication between these technical activities.

Among the great many advances achieved in recent years in electrotechnology, four major events have had major impact on the evolution and use of electronics data:

(1) The invention of the transistor and the subsequent development sequentially of whole families of low- and high-power semiconductor active devices, of integrated and hybrid microminiaturized circuits, of medium-scale (40-100 active elements) integrated arrays, and of large-scale (over 100 active elements) integrated arrays;

(2) The application of optical pumping to light energy to produce coherent focused electromagnetic energy -- the laser (light amplification through stimulated emission of radiation);

(3) The rapid development and application of digital techniques to computers and to high-speed communications; and

(4) The development of radically new power-generating or energy-conversion devices, such as radioisotopic thermoelectric generators, solar-cell arrays, and fuel cells.

The first -- semiconductors -- is of such significance that its effect has been felt by nearly everyone in the civilized world. By permitting engineers to cram more electronic functions into less space at lower power, instruments may now be built that either could not have been fabricated economically or would have been intolerably gargantuan before. The result can be seen from such opposites as super computers operated by only one or two people and the transistor radio pressed to the ear of a child strolling across the lawn.

The second -- the laser -- effected the marriage of electronics and optics, thus joining the use of entire electromagnetic spectrum and making light a useful energy source for the practical engineer.

The third -- digital techniques -- assures an ever-improving means to make better use of high speed data handling systems and provides a data transmission mode flexible enough to handle everything from inter-company facsimiles to deep-space communications.

The fourth -- new small power sources -- assures greater and greater flexibility and reliability in the long-term operation of instrumented craft or platforms, whether in space, on the ground, or on the ocean bottom.

It is from such progress as described above that electro-technology has permeated every field of engineering and every basic science. Today's engineering tools and scientific instruments all depend heavily on electronics.

The engineer uses instrumented satellites for improved geodetic data generation; automatic plotters for data display; electronics and electro-mechanics gear for operational-data monitoring and operation control; and computer simulations of aerospace systems long before a design reaches the fabrication stage.

The same advanced electronics data handling and sensing is used by the scientist, whether his efforts be for applied or basic research. The computer saves him time and electronic devices perform his measurements and analyses. The metallurgist uses X-ray machines for data generation; optical systems are designed by computerized modeling; the astronomer uses a radio telescope; and the biologist employs the electron microscope.

2. Data Characteristics

The answer to the problem of describing data used in all the various activities found today in the broad field of electronics/electrical engineering is in defining the scope of the field. The approach taken here is to select major, but different, parameters for the various use functions to show the variety and extent of data needs by system designers or engineers.

Subsystem Design. For the design of any electronic or electrical subsystem or for any major system, certain basic information or data must be known, since these data provide the stepping stones from which a final design may evolve Often, these basics determine whether a system will or will not ever reach the design stage. Unless funding or time is unlimited (and neither ever is), then a designer must determine, either from his own company files or from external sources, guidelines to the following:

(1) Reliability as expressed in expected lifetime operation and/or mean time between failures. Whether the system design is evolutionary or revolutionary, some guide in the form of hard numbers must be employed to assure or support a probability of success.

(2) Physical data for size, weight, and suspension characteristics are mandatory to assure that the system to be designed can, in fact, meet limiting parameters already established for the housing and/or transport of the system.

(3) Based on past experience, the designer must have some rapid means of determining power and either cooling or heating characteristics to assure that, for a given system, the use of auxiliary equipment will not obviate the overall system design due to the original physical limitations.

(4) Logistics with respect to maintenance and replacement parts must be considered, particularly for any system to be located or operated in a remote area -- a region where resupply may be difficult. Such data may relate also to items employed in the system which, by their very nature, are in short supply due to either material or manufacturing limitations.

(5) Finally, the designer must have some means of learning rapidly whether or not physical laws will deny him a successful system design. In many cases, while the "law" may be ill defined, it will, in fact, be described through available data. Thus, while a power engineer may know the load-bearing characteristics of his heavy power cables, he may need all the local characteristics of weather and terrain to verify long-term suspension of a cable over a valley. The radio engineer must know both the propogation characteristics and terrain effects of a particular area to verify his calculations for knife-edge defraction of radio waves over a mountain top.

The point is that every design is a compromise and, without firm reference data solutions to the same problems would be required endlessly. The relatively new field of microminiaturized electronic circuits encompasses, at least in part, the whole traditional field of electro-technology. For this reason, it is treated separately here. Through the use of tens or even hundreds of active semiconductor elements on a single small substrate, highly advanced subsystems are now being fabricated for all types of equipment in each field of endeavor. The microelectronic-circuit designer, whether his approach is through thin-film hybrid circuits or fully integrated monolithic circuits, must have data available covering both semiconductor material and basic-element physical characteristics plus the electrical characteristics of each. While much of the original data may originate with the physicist, chemist, or metallurgist, the engineer today must have such information as shown in Table II-B-1.

Table II-B-1. Characteristics for semiconductor-grade silicon.

Characteristics	Value	Remarks
Tensile Strength (no yield point below 600° C)	40-70X10^3 lbs/in.^{2}	30° C<T<600° C
Young's Modulus	25X10^6 lbs/in.^{2}	30° C<T<600° C
Thermal Conductivity	1.3-0.33 watts/cm-° C	300° K-1100° K
Specific Heat	0.14-0.22 cal./gram-° C	300° K-1100° K
Coefficient of Linear Thermal Expansion	2.5-4.8 ppm/° C	300° K-1100° K
Melting Point	1415° C	---
Dielectric Constant	11.7	---

For thin-film integrated circuits, the designer is concerned with the vapor deposition of very fine layers of conductors, dielectrics and insulating materials and the "paste-on" of active semiconductor elements. Thus, it is essential that he have available the characteristics of these materials to assist him in his design. Typical of the data required are those shown in the following Tables II-B-2, II-B-3, and II-B-4.

For the design of circuit elements, the engineer will find the need for a multitude of electrical comparisons normally presented graphically. For example, in the design of metal-oxide-semiconductor field-effect transistors (MDS FET), he must be concerned with insulate-gate field effects in monolithic silicon. He must obtain data comparing graphically the curves relating drain current (in milli-amperes) to drain-source voltage or drain current versus gate-source voltage.

Table II-B-2. Commonly used substrate materials with typical values of thermal constants.

Substrate Material	Thermal Conductivity (watts/in. -°C)	Coefficient of Linear Expansion (10^{-6}/°C)
Alumina (Al_2O_3) 96%	0.6	4-7
Beryllia (BeO) 99%	4.1	6-9
Borosilicate Glass	0.03	5
Pyroceram*	0.54	8
Quartz (SiO_2)	0.3	0.3

*TM Corning Glass Co.

Table II-B-3. Typical resistive materials for thin-film circuits.

Resistive Material	Sheet Resistivity (Ω/square)	Temperature Coefficient of Resistance (10^{-6}/°C)
Nichrome	200	± 50
Tantalum	100-1000	±200
Chromium	500-3000	± 50
Tin Oxide	to 500	±300
Cermet	to 20,000	±250

Table II-B-4. Dielectric materials commonly used in thin-film circuits.

Dielectric Material	Leakage (μamps/μfd)	Dielectric Constant	Dielectric Strength (volts/cm)
Silicon Dioxide (SiO_2)	---	~4	5×10^6
Silicon Monoxide (SiO)	<10	~6	3×10^6
Tantalum Oxide (Ta_2O_5)	$<10^{-1}$	~21	5×10^6

Greater importance for today's design of digital integrated circuits is expected. These commonly are employed in logic, memory, input/output, and power-supply functional circuits. Normally, they involve the problems of gating and temporary-storage circuitry interconnected with complex networks to manipulate digital signals in conformance with predetermined logical operations. Typically, then, the engineer may look at characteristics such as those shown in Table II-B-5.

Table II-B-5. Common integrated logic circuits: performance range.

Type Circuit	Propagation Delay (nanosecond)	Power Dissipation (milliwatt)	Fan-In	Fan-Out
Diode-Transistor Logic (DTL)	10-150	60-5	2-8	3-11
Transistor-Transistor Logic (TTL)	10-100	25-5	3-8	5-15
Direct-Coupled Transistor Logic (DCTL)	10-150	25-2	2-4	3-16
Current-Mode Logic (CML)	5-10	50-35	3-5	20-35

The heart of thin-film integrated circuit technology lies with the vacuum deposition of various compounds. These may be used for the coatings for resistive and conductive material or for capacitors and crossovers. Often the evaporation and deposition process from material to material is sequentially continuous. Thus, substrate materials to be used by the engineer become a prime factor in what may or may not be used for any given process under any given temperature extremes. Table II-B-6 below illustrates the characteristics that must be known by the engineer for standard substrate materials.

Table II-B-6. Standard substrate materials used in thin-film integrated circuits.

Parameter	Glass Borosilicate	Dense Alumina 94%	Dense Beryillia 98%	Sapphire
Softening Temperature °C	820.	1500.	1600.	2040.
Thermal Coefficient $10^{-6}/^{\circ}C$	3.25	6.2	6.1	6.0
Thermal Conductivity cal/cm/sec/°C at 25°C	0.0027	0.073	0.50	0.08
Density g/cm^3	2.23	3.58	2.90	3.98
Dielectric Constant	4.6	8.9	6.3	10.0

Communications systems. The transmission of information, whether in the form of voice or data, between men or between man and machine, has become a highly refined and well-documented engineering endeavor. Fundamental data are broadly available and equally well known to those in this field after nearly four-score years of development and implementation, particularly in the design of subsystems. Further, the field is highly regulated as to input and output characteristics, power and frequency limitations and man-generated electromagnetic interference by both international agreement and Federal agencies.

Systems application, however, often requires highly defined parameters for such operating modes as tropospheric-scatter, space relay of communications and both ground and space telemetry. For example, in the field of radio and television broadcasting, relatively standardized transmission systems can be readily designed and installed for any modulation or power output within legal limits. Yet, data must be known to achieve appropriate antenna design and deployment. Propagation anomalies and geologic magnetic influences bear directly on the type of radiator, its location and its height above ground. Such problems will be heightened with the introduction during the next five years of domestic TV relay via synchronous satellites across the United States. Both local and transcontinental propagation variances must be known for effective and efficient high-quality operation.

For long-range radio transmission, careful transmitter and receiver station location often is a factor requiring detailed knowledge of site Fresnel zones and atmospheric characteristics over the path length. For high-speed data transmission, the distribution of errors diurnally over a long-time span and long-term data dropouts anticipated due to atmospheric attenuation over the path length must be known. Predictions of periodic solar activity and the effects of such storms on signal strength for given Earth coordinates must be known to provide some means of estimating long-term system reliability.

The telemetry field, generally involving the remote measurement of physical variations and the transmission of such data to either manned or unmanned receiving stations, is similarly formalized and highly regulated. In addition to the Federal Communications Commission limitations placed on all communications activities in the U.S., the use of telemetry in Government systems is guided by the IRIG (Inter-Range Instrumentation Group) Steering Committee serving the Department of Defense and the Space Agency. Table II-B-7 reflects segments of a typical IRIG Telemetry Standard for system designs now in effect.

Table II-B-7. IRIG Standard FM/FM Telemeter Subcarrier Bands.

Channel	Frequency (Hz)			Frequency deviation (percent)	Maximum data frequency response* (Hz)
	Lower limit	Center frequency	Upper limit		
1	370	400	430	±7.5	6.0
2	518	560	602	±7.5	8.4
3	675	730	785	±7.5	11.0
16	37,000	40,000	43,000	±7.5	600.0
17	48,560	52,500	56,440	±7.5	790.0
18	64,750	70,000	75,250	±7.5	1,050.0

*Based on deviation ratio of five.

Electronic warfare. The problems of electronic warfare offer a unique problem for the system designer. Data required to assist him in his designs are available only to members of the inner "club" and he must show a "need-to-know" to obtain information from the Department of Defense.

In general, the field is related to that of electronic counter and counter-countermeasures (ECM or ECCM). It involves the jamming, confusion, penetration, or detection techniques as applied to the fields of radar, communications, and electronic guidance systems for missiles or aircraft.

The importance of electronic warfare cannot be underestimated in military systems, since it encompasses not only those systems employed to detect or confuse, but it involves all those systems whose essential function is other than electronic warfare. Thus, the designer of even conventional ground or airborne communications or radar systems must be cognizant of the radiating characteristics of his equipment with respect to enemy detection or jamming capabilities. Generally, however, the designer's responsibilities end with the development of his equipment and user elements adapt the operational and evasive tactics to go with the equipment.

The designer often must relate his power output density (in watts per MHz) to the power density that would be required from a broad-band jammer. Combined with an ability to change frequency, the system design that results most frequently is a compromise favoring the particular need.

The other side of the problem is that of finding the intelligence from within a staccato of noise, both natural and man-made. The designer must find a way to distinguish the desired intelligence from the noise by recognizing characteristic signal parameters: frequency, time of occurrence, signal duration and amplitude. Complex detection and correlation schemes may be required to ressurect a signal in the presence of determined jamming. While the designer may depend heavily on probability tables, much of his data may often only be obtained through extensive testing and the development of appropriate empirical data.

For reconnaissance or ferret equipment, the system designer is concerned with radio-listening search and analysis or recording. His search is concerned with the enemy electronic environment or any change noted therein. Thus, his equipment attempts to identify carrier frequency, pulse repetition rates, pulse widths, antenna scan rates and patterns and message characteristics such as modulation type or digital patterns. The designer's search for data must encompass the characteristics, or operating limits, of expected enemy electronic systems.

Electrical power generation. Like the field of communications, that of power generation is one of long history and heavy documentation. Whether the need be for low-or high-power design, data for conventional systems are available enmasse.

New fields and new techniques of power generation have created a different situation. One, in particular, is that of the use of thermonuclear or radioisotopic power generation, technologies that, until recently, were developed and controlled in large part by the Atomic Energy Commission and a few industrial firms.

The use of radioisotopic fuels, much like that of the microelectronic materials, requires the engineer to make heavy use of material characteristics. For example, in selecting strontium titanate as a basic generator fuel, the designer must know its density (g/cm^3), the melting point (^oC) and specific power (w/g).

Three other areas, with respect to necessary data, must be available to the designer: fuel availability and cost, and radiation safety figures. These might best be shown in Tables II-B-8, and II-B-9 below, extracted from data compiled by the Atomic Energy Commission.

Table II-B-8. Availability and Costs of Selected Radioisotopes*

Radio-isotope	Availability by year 1963	1967	1971	Useful Life (yrs.)	Projected costs ($/Weh‡)
	3 Mc	10	10		0.023
Sr^{90}	19 kwt**	63	63	10	
	0.75 kwe†	2.5	2.5		
	1 Mc	10	10		
Cs^{137}	5 kwt	48	48	10	0.031
	0.2 kwe	1.8	1.8		

*Electrical energy values assume 5% overall conversion efficiency and mission times shown.
**kwt = thermal kilowatts.
†kwe = electrical kilowatts.
‡weh = electrical watt hours.

Table II-B-9. Maximum permissible concentrations of radioisotopic fuels.

Radio-isotope	Fuel form	Occupational Exposure air (uc/ml)	Occupational Exposure water (µc/ml)	Nonoccupational exposure air (uc/ml)	Nonoccupational exposure water (µc/ml)
Sr^{90}	Soluble	3×10^{-10}	4×10^{-6}	1×10^{-11}	1×10^{-7}
	Insoluble	5×10^{-9}	1×10^{-3}	2×10^{-10}	4×10^{-5}
Cs^{137}	Soluble	6×10^{-8}	4×10^{-4}	2×10^{-9}	2×10^{-5}
	Insoluble	1×10^{-8}	1×10^{-3}	5×10^{-10}	4×10^{-5}

General characteristics. As indicated earlier in this section, the entire field of electro-technology has evolved into a highly-sophisticated, many-faceted industry whose data sources number in the thousands and whose data banks in total have reached astronomical proportions. In general, data for each major field of interest are available in summary form representing thousands of man-hours of long-term data reduction, analysis, and evaluation. The sheer weight of data for any given field precludes the storage of raw data and, in fact, modern data reducing techniques obviate the need for perpetual storage of raw information.

It is doubtful that any institution has a correct figure for even the data generated annually from electro-technological projects throughout the nation.

The economic value of such data is equally impossible to estimate. True, it has an intrinsic value which might be estimated as a percentage of development costs, but, like the foundation of a house, while its initial cost may have been relatively low, the entire industrial structure is supported by it. The rate of obsolescence is slow, since the technology is one that evolves and grows and today's advances are nearly always based on the shelf items and experience of yesterday. Even so-called scientific or engineering breakthroughs, which provide a step advance or produce an entirely new concept or product, are directly related to existing technology and success is based on the discovery of a missing link.

Proprietary considerations. A basic problem lies with the ownership or proprietary consideration associated with the bulk of basic data. In the past, the nation relied heavily on the universities for basic and applied research and on industry for both applied research and product development. Beginning in World War II, a serious change occurred. Heavy research and development funding became available from Government sources. The result was mass-produced research and development, with the Government retaining a larger portion of patent rights and, accordingly, demanding the turnover of the data developed to support each program.

Despite the end of World War II, Government-sponsored research continued, and beginning in the '50's, it introduced another change -- that of accelerated weapons development through concurrency. With the initiation of massive missile and space development programs, the nation experienced a new trend, still in existence today, wherein nearly three-fourths of all electro-technology development is funded by the U.S. Government.

The majority of Federal R&D funds were channelled into industry with the following result: universities found themselves largely funded for only pure research; so-called not-for-profit organizations arose in the form of large management-scientific-engineering complexes to fill a skilled-personnel void existing in the Government defense/space establishments; and industry, partly through its own initiative and partly through the need for filling technological gaps, greatly accelerated both basic and applied research to support the development of aerospace systems and techniques.

Thus, today we find sub-industries, such as those studying semiconductors, having accumulated and now controlling the bulk of the data in the field. In fact, these industries are years ahead of the leading universities in technological advancement. It is apparent that this trend will continue.

The field of microelectronic circuitry may be unique, but it is no accident that technologically it is led by one major research organization and five or six manufacturing firms which house most of the data. The reason is that each organization invested heavily in in-house research over fifteen years ago, accumulated the basic data needed for today's developments and with continued in-house sponsored research will probably continue to maintain leadership. They have developed and maintain their own data banks.

The same situation holds true in other fields. Despite the large number of computer and data processing firms in the United States, nearly 75% of U.S. sales and an even higher percentage of international sales are tightly held by one firm. Again, at least some of this strength can be related directly to a tight-fisted control of technical data.

In contrast, and largely due to its long history, we find the communications industry very broadly based if one excludes the near-monopoly held by one firm due to its function as operator of a national utility. All aspects of highly refined data in the communications field are available with few exceptions and these lie with highly specialized communication techniques applied to long-range military channels for strategic purposes.

There is an interesting comparison that can be drawn from two areas in the detection and tracking field, related at least in part to the dissemination of data. The sub-fields of radar and infrared were initiated, roughly, in the same time frame during World War II. Today, we find highly developed radar acquisition and tracking systems in use for both military and civil operations. The once highly classified radar data, following World War II, were made available to industry and, in large, were released from security classification for the general benefit of the nation. Conversely, the field of infrared technology still is cloaked under the mantle of military security and, except for very limited applications, it has been shielded from general civil use.

In the power field, it has already been stated herein that considerable empirical and statistical data are readily available in all branches with the exception of the use of nuclear energy. Data associated with the latter more and more are being released for industrial use, but the field is still clouded with misinformation and a shortage of data on hazards resulting with its use. Pressure by the U.S. Congress, this year, is expected not only to change this situation, but to result in the institution and enforcement of new data controls.

3. Data Flow

Data Generators. The users and the sources or generators of electrotechnological data are very often the same organizations. Typically, they include the universities, the research laboratories, the manufacturers, government agencies and departments, and, in some cases, centralized user organizations which control hardware development. These same organizations generate the data and, either directly or indirectly, through their associated technical societies or associations, disseminate information to others. (See Figure II-B-1.)

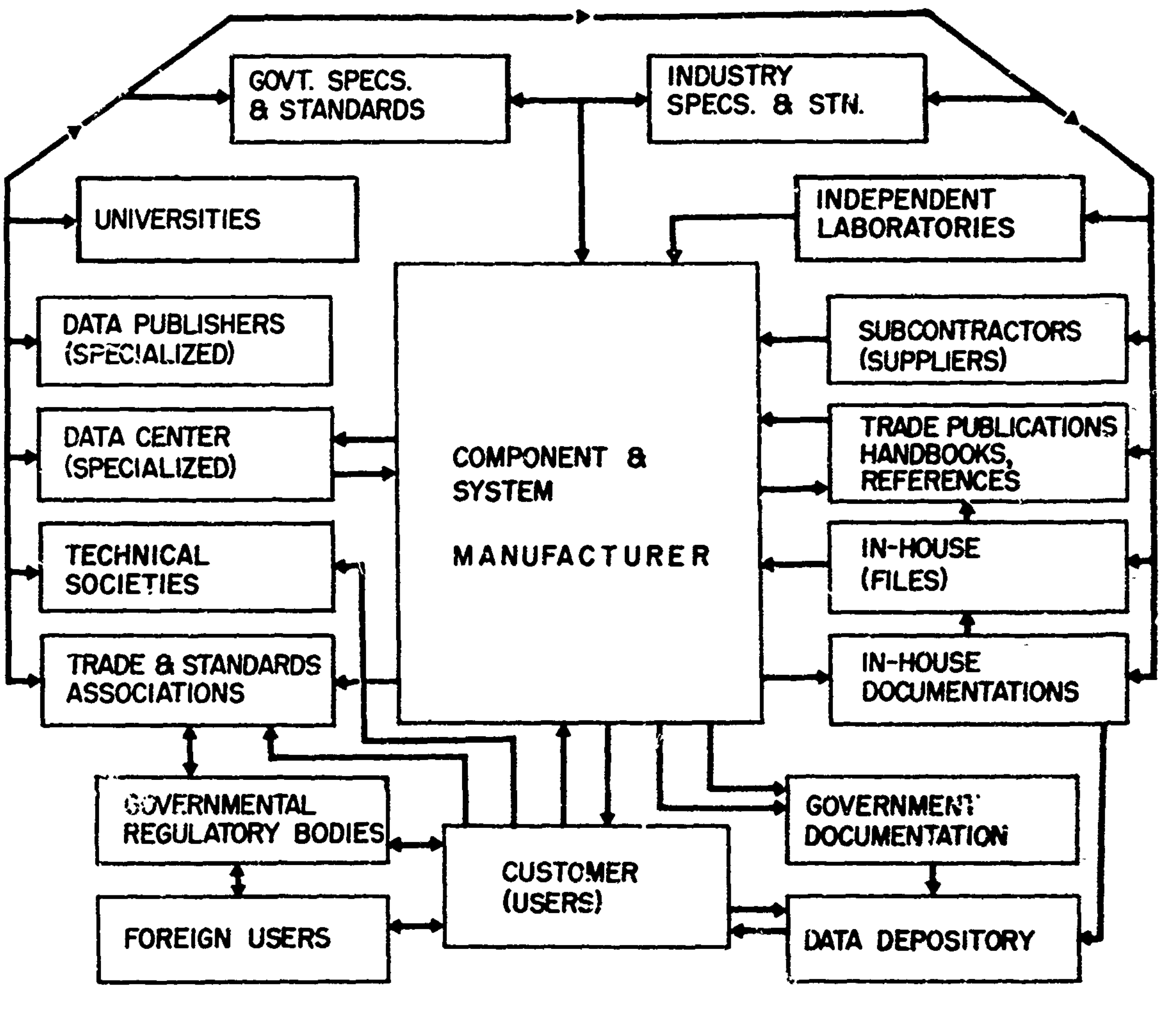


Figure II-B-1. Typical Data Flow Experienced in R&D Leading to Ultimate Production of Electronic or Electrical Components and Systems Produced Under Government Contract

All draw heavily on vendor-provided information for components, subsystems, and systems. Vendors freely provide catalogs showing electrical, mechanical and physical characteristics of every item manufactured. Also, they provide operational and performance information and, in many cases, can provide extensive backup with reliability information. The only data not provided openly by a vendor is that related to custom-developed hardware which might reveal either vendor or customer proprietary information and thus give a competitor either some technical or economic advantage.

The system developer is generally more restricted in the dissemination of data. A major system normally involves the integration of many subsystems and the application of operating techniques carefully designed to fit a particular requirement or mission. Thus, while the subsystems employed may be common in nature, the equipment complex may result in a proprietary or Government-classified configuration or operation.

Another form of data results today, not from a specific product or specific system, but from studies related to some complex operation or potential global network. Such information might be generated by a Government or private "think factory." Depending on the customer organization, such information may be broadly disseminated through the technical press, distributed on a limited basis through appropriate technical associations or societies, or may be buried under classification for decades.

The engineer, the physicist, the optical specialist, and the engineering manager in the field of electro-technology must maintain technical competence and must be thoroughly familiar with and use all the available sources of information within the field of his endeavor. In general, he relies on readily available published material. These fall, generally, into six categories:

(1) Textbooks or highly technical reference books abound for the entire field. Literally hundreds of new and revised hardbound and softbound books are published yearly covering every field of interest.

(2) Trade magazines and journals are available for the specialist and many are either free or require only token subscription fees. These provide not only general information concerning the entire field of electronics, but often provide highly technical treatises prepared by leading experts in each field. Generally, also, they have the advantage of providing timely information.

(3) Society papers and letters are available from technical conferences, symposia, and proceedings following meetings of such technical societies. A good percentage of these provide very specialized data on highly advanced technologies. In fact, the technical paper for a major society such as the Institute of Electrical and Electronics Engineers often represents the first public disclosure of a breakthrough or major advance in a field.

(4) Government reports and instruction books or manuals represent an excellent source of data on everything from the completion of major projects to the assembly, installation, and operation of major electronic systems. Except for classified programs, these are generally available either from the agency involved, through the Department of Commerce (a discussion of data-document depositories is provided below), or through the Government Printing Office in Washington, D.C.

(5) Industry reports, instruction books, and catalogs or data sheets are available upon request from all U.S. manufacturers.

(6) Government and industry standards and specifications are available and are generally disseminated by associated Government agencies, standards associations and technical societies. (One major exception is that of commercial aviation electronic specifications which are prepared and disseminated by Aeronautical Research, Inc., of Annapolis, Maryland, an independent firm owned by the U.S. airlines.)

Other standard sources of information regularly used by the professional engineer are the scientific and technical conferences and symposia requiring personal attendance. Some of these make use of invited papers and comprehensive panel discussions from which transcripts are made available only after a long time delay. It behooves the attendee to either make use of notes or personal contact with the speakers or panelists to obtain information needed for early use. A final and more informal source of information is by direct contact and establishment of rapport with specialists in the field at their places of employment.

For comprehensive data search and retrieval, more formal means may be applied by the engineer. He may, if working under Government contract, avail himself of the several National data centers now being established for various specific areas in electro-technology. Publishers of data have come into being providing information, either under Government contract or on a purely commercial basis. To assist in formal data collection, some data coordinators have been selected to direct the joint acquisition of information in given fields. Finally, limited data are available through several Federal Clearinghouses. The following typify present approaches employed for each of these data retrieval mechanisms.

The Electronic Properties Information Center (EPIC) has been established by Hughes Aircraft Co., Culver City, California, and is funded at about $250,000 per year by the U.S. Air Force Materials Laboratory. International in scope, the Center was started in 1961. Highly automated, it eventually will provide electrical/electronic materials properties for nine major categories. Initially, it is involved in the collection of data for semiconductor and insulator materials. The types of data available include direct-measurement electrical properties, energy-state measurements, and physical properties of crystalline structures.

NASA has located the Diode and Transistor Data Center at the Goddard Space Flight Center, Greenbelt, Maryland. Totally Government-funded, the Center was established several years ago and concentrates on U.S.-generated data only. Ultimately, it will provide comprehensive listings of electronic characteristics for twelve diode and transistor categories (by major functions). In addition, data will be related to U.S. manufacturers, procurement specifications and standards.

A third kind of center will evolve from the Systems Effectiveness Program now under way for Military Construction Facilities under the direction of the Advance Technology Branch of the Office of Army Engineers. Located near Washington, D.C., at Fort Belvoir, Va., it was established in 1963 to support the Nike-X (anti-ballistic missile) program, but will be moved in 1969 to the Army Construction Engineering Research Laboratory at the University of Illinois. The Center employs only limited automation. It is concerned primarily with the collection of data associated with electrical-power subsystems including engines (for primary power), generators, switch gear, transformers, etc. Types of data include those covering equipment performance, reliability, availability, and maintainability.

D.A.T.A., Inc., (Derivation and Tabulation Associations, Inc.), of Orange, New Jersey, is a commercial enterprise providing all types of data on semiconductors and integrated circuits under a Government-funded contract. Its effort draws on international vendor-supplied information covering each major component field and also provides appropriate military specifications.

On the pure commercial side is the Vendor Service Microfilm File (VSMF). This service is provided from the Microfilm Catalog File, Information Handling Services, Inc. Begun in 1960, it concentrates on aerospace and electronics data from U.S. manufacturers and can provide vendor catalog information, standards, parts and engineering drawings, and handbooks -- all available via microfilm.

The Autonetics Division of North American Rockwell, Inc., was recently selected to establish the Pacific Northwest Federal Agencies Data Management System at Anaheim, California. Acting as a data coordinator, the firm will develop plans for the joint collection and handling of all types of electrical-power data (the effort also includes the collection of hydrometeorological data).

A final source of limited data, as indicated above, are several of the Federal Clearinghouses which have been established for the collection and dissemination of documents. Typical of these is the Clearinghouse for Federal Scientific and Technical Information, Department of Commerce, located in Springfield, Va. It should be emphasized that data are available primarily through handbooks. Data listings for a given category cannot be obtained, but a listing of handbooks and the documents themselves are available from any of over fifty functional fields.

4. Data Management Problems

If one can summarize the problems of data management in the total field of electro-technology, then the situation might best be described as that of a totally u controlled information explosion. The industry is massive, diverse, and competitive. It employs parallel and often uncompromising technical societies which vie for recognition and professional leadership. The result is an undesirable but totally expected redundancy in data output. Several factors contribute heavily to multiplication of both data production and data management efforts.

Universities demand formal publication of technical and scientific data by their principal professors and graduate instructors. The result, often, is the publication of technical trivia and outmoded or inconsequential concepts.

Technical societies derive a portion of their income from successful trade shows accompanying technical conferences. To attract exhibitors, they must attract a high attendance by engineers. To attract engineers, they must present a braod technical program. The result is the preparation and delivery of papers on the same subjects and often with the same data over and over again throughout a given year.

The principal problem, then, is one of a total lack of selectivity on a national scale -- everything that looks technical can be and is published.

In contrast, some information that should be made available suffers from over-management by the Government. In the name of national security, the classification of whole blocks of information actually can create a void leading to the almost total absence of the civil application of a particular technology. It is relatively easy, with present Government procedures, to classify a subject to protect national security; although the mechanism exists for classification review and downgrading, the removal of a security classification occurs only infrequently.

Efforts to standardize terminology and to summarize and categorize information have been initiated for the various data banks across the country, but means must be found to apply more funding and more manpower to the rapid digestion and storage of appropriate data for automatic computer search, retrieval and printout. Without some means of making all the meaningful information on hand available in useful form within a reasonable period of time, the ever-growing mountain of data generated by the field of electro-technology will be largely non-recoverable and near-useless.

SELECTED BIBLIOGRAPHY

1. Parks, M.S.; The Story of Microelectronics ... First, Second, and Future Generations, Publication No. P3-75/317, Autonetics Div. of North American Aviation, Inc., Anaheim, Calif., 1966.

2. O'Donnell, C.F., et al; Applied Microelectronics, America House, Washington, D.C., 1966.

3. Stiltz, H.S., et al; Aerospace Telemetry, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1961.

4. Besserer, C.W., and Nixon, F.E.; Principles of Electronic Warfare, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1961.

5. Corliss, W.R., and Harvey, D.G.; Radioisotopic Power Generation, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1964.

6. Lynch, W.A., and Truxal, J.G.; Signals and Systems in Electrical Engineering, McGraw-Hill Book Co., Inc., New York, 1962.

C. Materials Science and Engineering

1. Introduction

For the purpose of this study, the field of materials science and engineering is defined as the study and use of knowledge concerning the structure and properties of solid materials, both metallic and non-metallic, particularly the information and data which pertain to the design and manufacture of products. The broader aspects of the structure and properties of solid, liquid, gaseous and plasma forms of matter are treated from a more theoretic standpoint in Section II-C to follow on Chemistry and Chemical Engineering. Some specific applications aspects of managing materials data are touched on in Section II-A on Aerospace Science and Technology and Section II-B on Electronics and Electrical Engineering.

The influence of materials data on almost every field of science and technology follows from the self-evident truth that materials comprise all the objects which are used in these fields. Materials sciences cover studies of the mechanical, physical, electrical, optical and chemical properties of materials, as well as fundamental structure and other characteristics that make materials attractive for specific uses and subject to specific treatments. The applications of materials to science and industry, to construction, to fabrication, and to engineering design cover the field of materials engineering. Materials engineering makes use of the sum total of data generated in materials sciences to design, develop and build optimal structures and artifacts. The uses and applications of materials depend almost entirely upon the data, generated and available, on the various properties.

It is therefore difficult for anyone to evaluate the significance of materials sciences and engineering data or their impact on the economy and well-being of the United States. A few statistics aid in comprehension of the tremendous importance of materials. In 1965: 2,754,476 short tons of aluminum were produced with a value of $1.349 billion; 90,432,000 short tons of pig iron were shipped at a value of $5.154 billion; 7,251,000 long tons of native sulfur were processed with a value of $145 million; and 2.843 billion barrels of crude oil were pumped with a value of $8.158 billion.

The significance of the materials sciences and engineering is shown by two widely accepted indicators of the economic health of the United States. These are the annual rate of private housing starts and new automobile sales. By the last quarter of 1967, housing had reached an annual rate of over 1,500,000 private starts and auto sales totalled 8,400,000. These two industries alone consume tremendous quantities of materials from our forests, quarries, mines, steel and rubber factories, and synthetic fibers and plastics.

As of late 1967, over 200 million U.S. inhabitants are involved in the generation and/or use of data on the properties of materials. Every professional man (using the broad connotation of professional) works with materials. Members of the engineering profession, in particular, are important users of materials data. A few figures on membership of the founder societies illustrate this interest. The membership of the American Society of Civil Engineers was 59,444 on September 30, 1967. Of this number, 22,692 belonged to the Structural Division; this is almost 38% of the membership. The American Institute of Chemical Engineers has a membership of 31,515. A recent limited survey of the members shows that about 70% deal directly with materials in extraction, processing, production, design, safety or maintenance. A separate Division of Materials Engineering and Sciences is being organized by this Institute with a five day international conference and exposition, wholly devoted to materials, scheduled for early 1968.

The Institute of Electrical and Electronic Engineers has the largest membership of the founder societies in engineering, with a total of 175,000 members. These engineers depend heavily on materials for all of their devices, equipments, and electronic components. There is no record of any separate division devoted to materials. Other sources of materials data are utilized by electrical and electronic engineers. The American Society of Mechanical Engineers has a membership of 60,000. Most of this group of engineers are vitally concerned with materials in the generation of data, the use of these data in design, and in the fabrication and operation of designed equipment and structures. The American Institute of Mining, Metallurgical and Petroleum Engineers has 43,705 members whose interests range from fundamental studies on the solid state structures of materials to extraction of metals from ores and production of crude petroleum and natural gas. It is self-evident that no member of this institute can function without great dependence on materials and the data covering all of their diverse properties.

There are no lines of sharp demarcation among the many relevant disciplines involved in the overall activities of materials sciences and engineering. However, one can categorize the disciplines generally into materials sciences (under which fundamental property data are generated and new materials developed), materials engineering (under which materials data are used for design and for fabrication of structures, equipments and artifacts), and into materials users and consumers (under which devices, artifacts, and equipments are produced and utilized in processing both hard and soft goods for the ultimate end-user). Figure II-C-1 is a sketch depicting some of the relationships and lines of communication between relevant disciplines. It must be noted that only some disciplines are shown and that the lines indicate materials data flow to and from any selected disciplines.

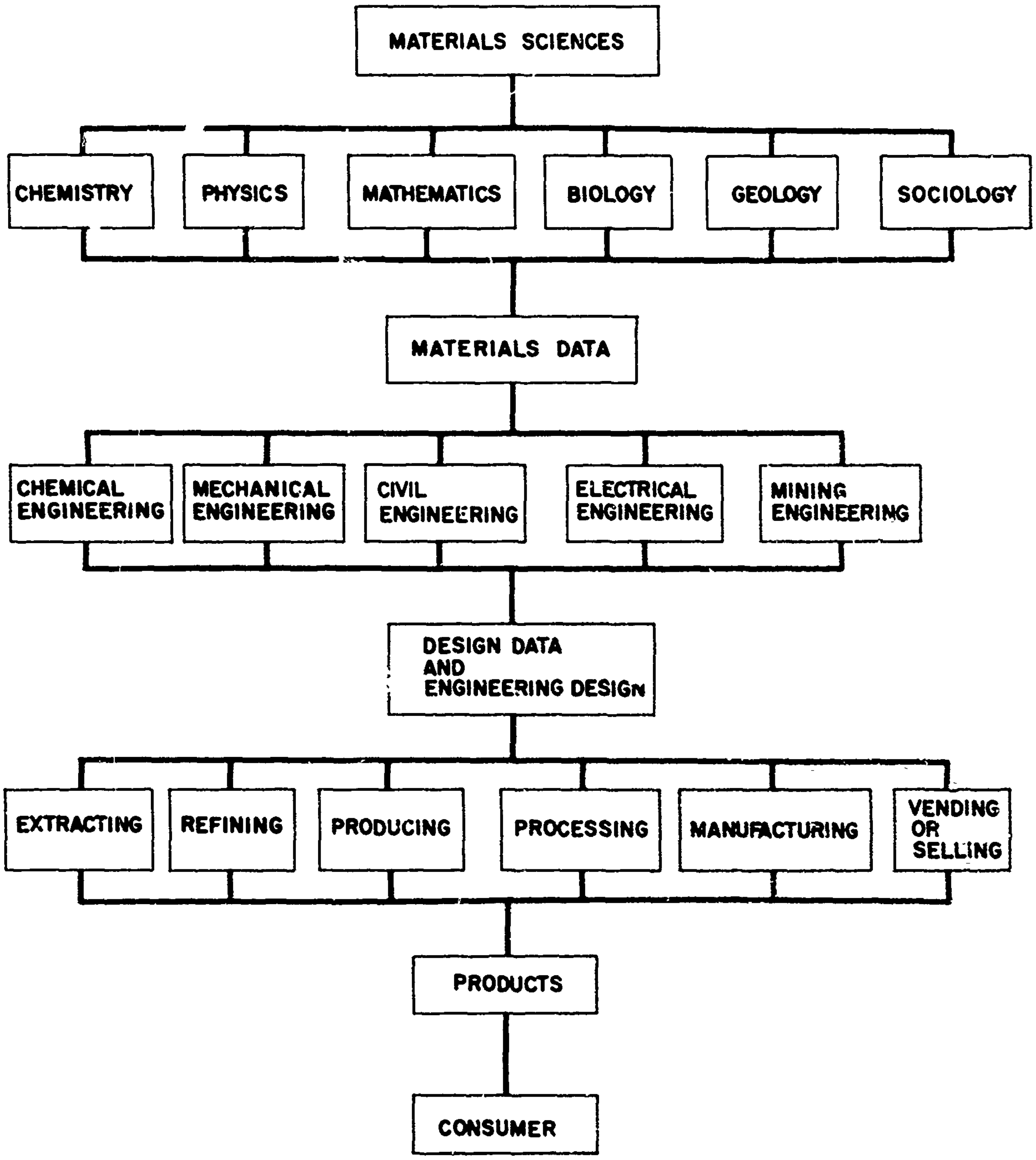


Figure II-C-1 Relationships of Some Relevant Disciplines

2. Data Characteristics

For this study, materials are defined as those things that are solid at ambient conditions. Data for materials sciences and engineering cover measurements of properties that can be reported in quantitative units. As Lord Kelvin stated many years ago: "When you can measure what you are speaking about and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind." An extension of Lord Kelvin's statement leads to an emphasis on measurements expressed in units that have exact meanings for each and every user of the recorded data.

Historically, the wealth of collected data has been classified along lines of production and usage interests. This means that communities of interest were formed around classes of materials, and data generation and dissemination were founded on the material classes as the common unifying element. This emphasis on a common interest in a given material may have arisen from economic considerations; e. g., International Nickel Company derives little economic advantage from promotion of concrete or glass materials. At the same time, advances in technology moved along materials availability, as evidenced by the "Stone Age," the "Bronze Age," the "Iron Age," and now the "Plastics Age." If materials are competitive - and in the real world, they are - then logical, economic survival depends on promotion of properties that have competitive advantages, by those producers, companies, or vendors whose economic existence depends on profitable sales.

For purposes of data generation, collection and classification, materials data are divided into two major groupings which evolved from classification of materials themselves: metals and non-metals. Table II-C-1 lists the initial classes of metals for direct consideration. It is more difficult to make a classification of non-metals. Table II-C-2 furnishes a practical breakdown of non-metallic materials data.

TABLE II-C-1. CLASSES OF METALLIC MATERIALS DATA

1. Cast and wrought iron
2. Mild steels
3. Alloy steels
4. Stainless steels
5. Copper and copper alloys
6. Nickel and nickel alloys
7. Aluminum and aluminum alloys
8. Magnesium and magnesium alloys
9. Zinc and zinc alloys
10. Lead and lead alloys
11. Titanium
12. Tungsten
13. Beryllium
14. Liquid metals - mercury, sodium, potassium, lithium, calcium
15. Noble metals - gold, silver, platinum
16. Tantalum
17. Molybdenum
18. Rare earths - cerium, thorium
19. Heavy metals - radium, uranium, bismuth
20. Specialty metals - cobalt, columbium, zirconium

TABLE II-C-2. CLASSES OF NON-METALLIC MATERIALS DATA

1. Wood and cellulosic materials

2. Ceramics
   a. Glass
   b. Stone
   c. Silica
   d. Porcelain
   e. Stoneware - terra cotta, pottery
   f. Refractories - acid, basic, neutral
   g. Building brick and firebrick

3. Polymers and plastics
   a. Thermoset
   b. Thermoplastic
   c. Reinforcements and fillers

4. Elastomers
   a. Natural
   b. Synthetic

5. Fibers
   a. Natural
   b. Synthetic

6. Concrete - mortar, plaster, lime

7. Leather - furs, skins

8. Cork - seals, insulations

9. Carbon and graphite

10. Asphalt - pitch, tar, bitumens

11. Specialty non-metals
    a. Carbides and nitrides
    b. Gums and waxes
    c. Solid lubricants - molybdenum sulfide, lead sulfide

Materials data are also categorized according to various classes of properties. This method of classification was uncommon until recent years. Its use has been increased by the multiplicity of new materials, developed to achieve specific properties for special end uses. An example of this is the use of composited plastic materials as heat shields on space vehicles for re-entry, fo. which ablative properties are the important characteristics for these unique end-use materials. Various government facilities have encouraged the classification of materials data by properties with their support of property oriented data centers. Selected classes of materials property data are listed in Table II-C-3. This listing is not exclusive. Omission of any class does not presuppose its lack of importance. The significance of any data class is dependent upon the needs of the particular user. As an example, a design engineer for highway construction is highly interested in the mechanical property of compressive strength and the physi l property of abrasion resistance, but has no need for the acoustical property of noise resistance or the nuclear property of radiation transmission.

Another technique for classifying materials data is based on use of fabrication categories. This technique leads to grouping materials under types of fabrication. In many cases, metals and plastics fall into the same classes such as casting, extrusion, molding and rolling. Concrete and cast iron also join in a class of casting compounds, while reinforced plastics and many metals and alloys have pertinent machinability properties. The use of adhesives for building structures leads to an adhesion property that is almost universal in its applicability to laminated papers and plywood, to reinforced plastics and to bonding of aircraft structures. This technique has advantages in providing some simplicity for fabricators, as does the property of machinability, but ceases to be universally acceptable for all materials.

TABLE II-C-3. SELECTED CLASSES OF MATERIALS PROPERTY DATA

1. Mechanical
   a. Tension
   b. Compression
   c. Flexure
   d. Shear
   e. Bearing
   f. Creep and Creep Rupture
   g. Fatigue
   h. Elasticity

2. Physical
   a. Density
   b. Hardness
   c. Melting and Boiling Point
   d. Color and Odor
   e. Water and Solvent Absorption
   f. Flow and Viscosity
   g. Solubility

3. Chemical
   a. Reactivity
   b. Corrosion
   c. Fire Resistance
   d. Environmental Resistance
   e. Chemical Structure
   f. Chemical Bonding
   g. Absorptivity and Adsorptivity
   h. Catalysis

4. Electrical
   a. Resistivity
   b. Conductivity
   c. Dielectric
   d. Magnetic
   e. Piezo-electric

5. Thermal
   a. Degradation
   b. Conductivity
   c. Expansion
   d. Diffusivity
   e. Ablation
   f. Emissivity

6. Optical
   a. Reflectivity
   b. Transmission
   c. Selectivity and Filtering
   d. Color
   e. Refraction

7. Nuclear
   a. Permeability
   b. Shielding
   c. Degradation
   d. Absorption
   e. Radioactivity
   f. Decay
   g. Chemical - change from Radiation

8. Acoustical
   a. Sound Transmission
   b. Vibration Dampening
   c. Sonic Degradation
   d. Audibility

It has been suggested that all materials data might be classified according to end-uses or users. Thus, all materials that have properties suitable for highway construction are classified as highway materials; those materials conventionally used for homes are categorized as home building materials. This categorization is needlessly limited and confining. For example, where does one classify the ordinary red brick which is used for pavements, for home building, for stacks, for garden walls and for industrial construction? Aluminum is another versatile material, whose uses range from cooking utensils through wall panels to skins of aircraft, so that a whole landscape of potential uses and users becomes involved.

It is easy to illustrate the vast complexity of the various data efforts on materials sciences and engineering by choosing the subject of "Polymers and Plastics." If one analyzes this subject, one determines that plastics come from polymers with thermoplastic or thermosetting properties. Both types of plastic materials can be used with fillers or reinforcements to obtain better strengths. At this point, keywords in plastics are "polymers," "fillers," "reinforcements," "thermoplastic," "thermosetting." In further analysis, polymers are formed from monomers; polymerization is controlled by temperature, pressure and the presence frequently of hardeners, catalysts or curing agents. New keywords have now been added: "monomers," "polymerization," "temperature," "pressure," "hardeners," "catalysts," "curing agents." So the first thing on any material is to analyze the system for all of the subheadings and keywords.

This discussion illustrates the many configurations that data on "Polymers and Plastics" as materials may have. Each user or potential user needs to choose a body of descriptors that fit his ultimate desire and purpose in order to select a method of obtaining the property data of immediate significance to him. Any data classification system ultimately must consider the needs of the potential users. This statement is of increasing importance as one passes through the threshhold of old-time conventional materials into the world of new-time, non-conventional and tailor-made materials.

Data Characteristics. Materials sciences and engineering data have certain common characteristics, on which usage of the data depends. These are: visibility, accessibility or availability, and viability. Data hidden by industrial or government secrecy or not published by researchers become invisible to most potential users and, therefore, do little to improve the usage patterns. Even when data are visible, they must be made accessible and available to every scientist and engineer. It is not sufficient to know that data on a material are generated and stored in some industry, university or government facility. The question becomes, how does one acquire visible data? Finally, materials data must be viable; i.e., accurate and reproducible, not obsolete, and useful to the using scientist and engineer. Presumably, viability is a criterion that is frequently overlooked. Any materials data activity that fails to provide visibility, accessibility, and viability cannot be termed significant.

Materials data range from the raw condition, as they are generated under laboratory and test conditions, to the highly sophisticated and evaluated data, as they are presented by competent authorities in handbooks. The degree of refinement required by the user depends on the usage to which the data are to be put. Most researchers in the materials area prefer raw data collected from publications of their peers or generated by their own activities. These raw data are used by the researchers to evaluate, compare, and refine their own generated data so that conclusions are better validated. On the other hand, design engineers demand evaluated, refined, and valid materials data so that buildings, highways, equipments, and other structural elements will serve their useful function without failure and with minimum maintenance requirements. The latter criterion of low maintenance with consequent greater economy is of major importance to the final user. Safety, health, and product purity are also valuable considerations by the ultimate user.

The volume of materials data, both available and unavailable, is enormous; it has not, and probably cannot, be measured. One problem that would immediately complicate any measurement of volume is duplication and replication of materials data. However, the following examples furnish some idea of the tremendous flow and volume of materials data. Over 100,000 references and over 10,000 separate data-containing items were collected during the preparation of the first edition of a "Materials Design Handbook Division I Structural Plastics" by Grove and Pray. Data were limited to certain types of reinforced plastic composites suitable for aerospace vehicles.

The Mechanical Properties Data Center (Technical Information Systems Division, Belfour Stulen, Inc., January, 1967) states that the "file contains more than 600,000 individual material test records. These include test procedures and mechanical properties of approximately 4,000 metal alloys. More than 8,000 new test records are added to this file each month. It has been conservatively estimated that this file represents the results of approximately $60 million in materials test programs." The activities of this data center are limited specifically to the mechanical properties of selected metal alloys of pertinent value to the Department of Defense.

It is obvious that the economic value of materials data is incalcuable. It is apparent that the civilized world would cease to exist if all accumulated materials data, written or otherwise stored, were lost by some catastrophe, such as a nuclear holocaust. Some indicators are available in confirmation of these broad generalizations. Cahners Publishing Company publishes a broad line of trade journals covering materials data and uses for consumers. In 1966, 21 journals were being published. During that year, Cahners merged with the International Publishing Corporation of Great Britain which was publishing and distributing 82 consumer trade journals largely devoted to materials. This merger gave a combined capitalization of over $372 million with annual income of over $36 million. Plastics World (1966 circulation about 44,000) and Reinforced Plastics (1966 circulation about 11,000) are two Cahners trade journals. About 50% of the space

in each issue of these journals is devoted to meaningful data on plastics materials. Less than 20% of the data available are published; the balance are stored and can be made accessible through Reader Service Cards (Cahners Publishing Co., Inc., 221 Columbus Avenue, Boston, Massachusetts (02116)).

Another important indicator of the economic value of materials data is the size of the publishing efforts of a scientific and engineering major publishing firm (John Wiley and Sons, Inc , 605 Third Avenue, New York, New York). Interscience Publishers handles the majority of the books on science in the John Wiley and Sons, Inc., organization. The engineering books are mostly produced by the Wiley Division. In each portion of the company, the majority of published material comes from textbooks; in 1965, approximately 50% of the sales came from textbooks, 35% from monographs and reference books, 10% from encyclopedias, and 5% from journals. In 1965, this organization published 376 new books. Of this number, 165 were considered undergraduate college texts; 28, graduate college texts; 118, monographs and reference books; and 65, imported scientific books. Eliminating the imported books from a percentage breakdown, about 25% of the books published would be considered materials data resources books. Wiley differs from some other publishing houses in their distribution; e.g., Academic Press handles about 80% of books that might be considered monographs and reference books and only about 20% texts. Wiley has very few, if any, textbooks for distribution to high school and grammar school students.

Materials data do become obsolete. However, the rate of obsolescence varies from material to material and is a function of the "newness" of the material. Older conventional materials, such as bronze, cast iron, and mild steels, have been used for many years as structural materials; the test methods are well standardized; the property data are well organized and readily available; designers and end-users alike have complete confidence in the data on any property or combination of properties. Data on the older, conventional materials seldom become obsolete.

On the other hand, new and less conventional materials are being developed for potentially exotic uses, such as composites of boron fibers with new refractory resin matrices. The end-uses for these materials are initially poorly defined because the environmental conditions of outer space to which the composites will be exposed are presently relatively unknown. Test methods are under development, which impose temperatures of -400° F. and pressures of a nearly perfect vacuum; manufacturing methods are being developed; and property data are either non-existent, poorly organized, or highly classified under National Security Acts. Change of materials and test methods makes property data rapidly obsolescent; confidence limits are unknown; and users believe it necessary to develop their own laboratory and design test data. In such a state of flux, presentation of property data in a formalized manner becomes a worthless objective.

It has been pointed out earlier that most property data for materials sciences and engineering are specifically oriented towards a given material. Even those data centers which purport to be property-oriented are somewhat directed towards a given type of material. For example, the Mechanical Properties Data Center (referred to later) has data available for dissemination only on metallic alloys. Some industrial associations also direct their data efforts towards a material orientation. Typical is the Technical Data Center of the Copper Development Association, Inc. (CDA Technical Data Center, Batelle Memorial Institute, 505 King Avenue, Columbus, Ohio 43201). More details of data orientation will be presented in the subsection to follow.

3. Data Flow

Data Users. In a very real sense, almost everybody is a user of materials data. At times, the user may not realize his dependency on materials and on the properties that distinguish one material from another. A child may question the bursting of his rubber balloon, but does not understand the property of elasticity. A housewife worries about the rusting of her iron skillet, but is not interested in the property of iron corrosion. Every maintenance man has a need for materials data, although in many cases, his need is satisfied by acquired practical data. This study is primarily concerned with two classes of users who, in general terms, recognize the necessity of materials data and consciously seek such data pertinent to their work. These are materials scientists and engineers.

It is difficult, in the area of materials sciences and engineering, to isolate data users and to classify them. If one considers any given material, the path of data generation and use follows some logical line, with feed-back from each successive user. The initial user of property data for any given material is the material scientist. Properties (see Table II-C-3) will be determined and used in further development of newer properties, newer materials, or potentially newer materials uses. In the grouping of materials scientists, one includes chemists, physicists, mathematicians, biologists, geologists, and even sociologists interested in the impact of the material and its uses on mankind.

The next major class of materials data users is the engineering community (the whole spectrum of engineers). This class of users requires data for design of structures, buildings, equipments, highways, armaments, and many different artifacts. The design may be as simple as a child's toy or as sophisticated as a space-probe rocket. The engineer prefers materials data that are evaluated, refined, and ready for use, under known confidence limits. Obviously, the engineer may be oriented towards research and development or, at the other end of the scale, towards practical operation of all devices and artifacts.

Working with the design engineer is a broad spectrum of materials data users. These people are those responsible for extracting metals or non-metals, for refining the particular material, for forming or fabricating the substance into things, for manufacturing articles and artifacts, for production items and equipments, and for vending or selling the products. After the product is sold, there arises a group of distributors, technical salesmen, transporters, and other intermediaries, who need and use materials data. Thus, in the whole proceedings of starting with a raw material and ending with a product user, there is a broad range of persons who require materials data. These are the supervising personnel in extracting, refining, producing, processing and manufacturing. Complementing the supervision, on one side, are management personnel, and on the other side are operating personnel who need data.

There are many supplementary materials data users who need a somewhat superficial knowledge of materials. In a manufacturing operation, there are maintenance people: boiler-makers, carpenters, electricians, instrument repair men, insulators, laborers and helpers, machinists, painters, pipe fitters, riggers, sheet metal men, welders, and others. When there's something to repair, construct, or maintain, they go where the job is located. There are additional personnel, whose interest in and need for materials property data may be considered on the fringe. These are: purchasing agents; safety men and industrial hygienists; claims agents for industry, commerce, and insurance carriers; industrial compensation judges and courts; attorneys for plaintiffs and defendants in many types of civil and criminal actions; and a whole variety of salesmen of retail products and household services.

The materials scientists, as users of data, are mainly concerned with basic properties and are willing to accept and often prefer raw, unrefined data taken from the original research reports or publications. For example, the chemist is concerned with basic chemical phenomena - reactivity, corrosion, chemical structure, and bonding. (See Table II-C-3, Item 3.) The physicist is more likely to be interested in solid state structure or in electrical properties, such as dielectric or piezo-electric effects. The mechanical engineer is definitely involved in using mechanical data on materials for design of machinery and for fabrication techniques, while the civil engineer is interested in similar property data for buildings, highways, dams, and reservoirs and for various other public works activities.

The complexity of the user population is illustrated by analyzing data use for single materials such as iron. Some form of iron oxide is the basic ore for the production of iron. Pig iron production in 1965 totalled over 90 million tons (loc. cit.) which, if extracted from relatively pure oxide ores, means over 125 million tons of iron ore had to be processed. The geologist is primarily interested in locating iron ore bodies and in the properties of the ore. Some of these properties are chemical (assay percentage); others are physical (friability, size distribution); others are electrical (magnetic for separation). Now, the mining engineer becomes involved; he needs various types of data to perform steps such as crushing, screening, grinding, concentration (electrostatic, flotation, magnetic, air cyclones),

agglomeration (compacting, briquetting, pelletizing, sintering), and heat hardening to prepare the iron oxide for the smelting operation. The metallurgical engineer usually takes over at the smelter plant and follows the iron through the blast furnace operation; he is vitally interested in chemical purity data, fluxing temperatures of slag, quantity and heating value of blast furnace gas, quality of coke and its production. The final product of this process of operations is pig iron. Table II-C-4 is a simplified listing of primary materials data users concerned only with production of pig iron, starting with the iron ore mining operation. Table II-C-5 lists supplemental materials data users in the same process. Inspection of these two simplified tables leads to the irrefutable conclusion that in any processing or manufacturing operation, everyone from top management to the lowest laborer is vitally interested in valid materials data on properties, as well as other fringe materials data of direct economic import.

The complexity of materials property data needs is demonstrated by the consulting engineer user. At a given moment, he needs mechanical strength data on a storage tank for natural gas, as well as the explosive limits of the gas and air mixture; he is concerned with the force exerted in any explosion and the possible source of ignition. At another time, he is involved in determining the velocity of a vehicle from the mechanical properties of its materials and the physical measurements of the inflicted damage.

The Materials Advisory Board of the National Research Council has used a more gross breakdown of materials users for the study: (1) Basic research; (2) Applied research; (3) Design engineering; (4) Administration; (5) Information activities; and (6) Service activities. Users were interviewed from industry, universities, government, and research institutes. The study was intended mainly to find out what materials users think about dissemination of information on materials by centers supported by the Department of Defense. The people interviewed feel that the Department of Defense is "doing all that it should in dissemination of information. However, the survey results indicate quite clearly that the centers and collections and their coverage, services, and accessibility should be publicized more effectively to reach a greater number of those who have need for special information services."

TABLE II-C-4. PRIMARY MATERIALS DATA USERS IN PRODUCTION OF PIG IRON AND TYPES OF DATA USED

1. Geologist - Exploration of Iron Ore Body
   a. Chemical
      (1) Composition of Ore
      (2) Ore Structure
      (3) Reactivity and Smelting
   b. Physical
      (1) Density
      (2) Hardness
      (3) Friability and Size Distribution
   c. Mechanical
      Bearing
   d. Electrical
      Magnetic

2. Mining Engineer
   a. Chemical
      (1) Composition of Ore
      (2) Smelting
   b. Physical
      (1) Density
      (2) Melting Point
      (3) Grindability and Compactibility
   c. Thermal
      (1) Heat Flux and Flow
      (2) Degradation
      (3) Sintering
      (4) Heat Hardenability

3. Metallurgist
   a. Chemical
      (1) Composition of Ore
      (2) Composition of Pig Iron
      (3) Reduction of Ore
      (4) Composition of Flux Materials

b. Physical
   (1) Viscosity and Flow of Pig Iron
   (2) Viscosity and Flow of Slag
   (3) Compactibility of Beneficiated Ore
   (4) Size and Compressive Strength of Coke
   (5) Blast Furnace Gas Flow

c. Thermal
   (1) Fluxing of Ore, Coke, and Limestone
   (2) Heat Flow in Blast Furnace
   (3) Expansion and Contraction of Furnace
   (4) Expansion and Contraction of Charge
   (5) Heat Balance Considerations

TABLE II-C-5. SUPPLEMENTAL MATERIALS DATA USERS IN PRODUCTION OF PIG IRON AND TYPES OF DATA USED

1. Management
   a. Chemical Purity
   b. Physical Size and Shape of Product
   c. Economics of Sales and Profits

2. Supervision
   a. Chemical Purity
   b. Physical Flow and Casting of Pig Iron
   c. Economics of Labor Utilization
   d. Economics of Coke and Gas Production

3. Technical Salesmen
   a. Chemical Composition
   b. Production Costs and Sales Profits
   c. Production and Delivery Schedules

4. Maintenance Personnel
   a. Chemical on All Materials
   b. Physical on All Materials
   c. Mechanical on All Materials
   d. Electrical on Some Materials
   e. Thermal on Some Materials

5. Safety Men and Industrial Hygienists
   a. Mechanical and Physical on Many Materials
   b. Toxicity of All Materials

6. Purchasing Agents
   a. Property Data on Many Materials
   b. Economic Data on Many Materials

Data Sources and Generators. Definitive treatment of the sources and generators of materials data is difficult due to the great numbers of materials and materials properties. However, it is possible to generalize on materials data sources and generators by identifying them into seven definitive categories:

(1) Personal files (consultants, academic and/or industrial personnel);

(2) Company or industrial files (staff responsibility through technical library resources);

(3) Technical book publishers (editorial staff, review and advisory boards, market research and sales personnel);

(4) Trade associations (limited scope of reports, data dissemination confined to participating member industries);

(5) Trade journals (limited volume of published data, additional data available on request, broad variety of readers, good coverage of industry);

(6) Professional societies (discipline-oriented, good to excellent review before publication of data-containing articles, editorial policies variable, journals of variable quality); and

(7) Data centers (specially oriented, collected data evaluated, rigid format, limited circulation and dissemination).

The first category, "personal files," includes the individual data resources consisting of miscellaneous or specialized textbooks, monographs, and handbooks, on which the individual depends for data necessary to carry out his assignments. The books are augmented by personal research notes, unpublished bulletins and theses, confidential memoranda and reports, as well as collected abstracts, reprints, and correspondence. The collection has maximum usefulness only to the particular individual user. The breakdown of

personal files varies widely and depends on the professional discipline and the nature of the individual's occupational use.

An academician's primary data source may be largely published articles, augmented by personal notes, theses, and reports, but all of which is subject to publication. A major use of these files arises from the teaching-research habits of the individual. Consultants, on the other hand, have not only book collections, mainly handbooks, but rely heavily on personal notes, reports of prior studies and investigations, and on other data that have not been and probably will not be published because they are proprietary to clients.

Similar data resources are available to the industrial scientist, engineer, and technologist, but are depended upon to a lesser degree because of the availability of company files with their rather strict company-confidential classiciation. Subject to the problems engendered by use of a limited sample, the personal files of an average, but experienced, person contain about 30% unpublished data and about 70% published data. These files are probably not useful to other scientists and technologists, except for those few people who are most closely associated with the individual. The two major advantages of personal files are: (1) The data are collected on specialized subjects, and (2) The data are organized for ease of retrieval and use by the person involved.

It is not easy to comprehend the sum total of materials data, stored or "lost" in the personal files of individuals. Some conception of the magnitude and value of individual files is visible from the fact that the five founder engineering societies have a combined membership of 369, 664 persons as of December 31. 1967. It is a reasonable deduction that 50% of these engineers deal directly with materials data in processing, developing, building, operating, design, testing, and other functional operations. The members of any technical society are the more professional and more productive users of materials data. One decides, therefore, that massive data resources are stored, but not broadly available or visible. Even if the duplication of data is considered, the questions become vital - Should, and if so, how should, these resources be made more readily visible and available to a larger number of users?

The second category of data sources, "company or industrial files," is defined as the sum total of data and information collected and stored by all of the present and past employees of the company or industry for the use of the present and future employees. The complexity of the company data files varies according to the size of the company. Large companies maintain extensive library holdings, including many journals and books, as well as company reports and memoranda. Company-generated data are usually proprietary and are not available generally for use outside of the company or its licensed affiliates. Company employees depend heavily on the "company files", and seldom do they compile extensive files of their own. In larger companies, the data files are serviced by technical library personnel, supervised by a trained scientist or technologist. At the duPont Experimental Station, information assistants and information systems are used to relieve senior scientists of arduous data searches.

It is important to note that materials data are generated by two major types of industries. One type is the producer and vendor of primary materials, which is generally desirous of increasing the use of each produced material. In the metallic materials, the data generated by International Nickel, Aluminum Company of America, and many other producers of metals are widely distributed to all potential users. Inquiries for data are promptly answered; handbooks are available for any user on request. Similar visibility and availability of data is noted for the producers of non-metallic materials, such as Shell Chemical, with their range of epoxy resins, and Owens-Corning, with their variety of reinforcements. The other major type of data generator in industry is the manufacturer or producer of finished end-products. Data on properties of materials used in the end items are highly proprietary and are only infrequently made available even to the suppliers of primary materials. Thus, materials data generated by manufacturers and fabricators of end-products are seldom visible or available, and these data resources can be categorized, but cannot be readily identified or enumerated.

Data generated by government contractors belong in the public domain within the limitations of security of the nation. These data are made visible through documents and reports, as well as by other means, such as operating drawings and manuals. They are thus available and accessible to those who have a "need-to-know."

Some data on end-products are exposed through papers presented at technical conferences and professional meetings; however, these are limited and of doubtful value, as the papers are frequently given solely in order to increase sales potential. Search of the literature to locate these data sources is difficult, time-consuming, and of doubtful value. This problem appears to be one which is inherent in obtaining data from end-product personnel, as compared with the relative ease with which data could be acquired from suppliers of materials.

The third of seven categories of data sources listed is technical book publishers, which are defined as publishing companies that specialize in monographs, reference books, handbooks, encyclopedias, and textbooks written for the use of materials scientists and technologists. Competent and experienced authors of these books are solicited by representatives of the publisher to write or edit for a given group of readers. Commercial publishers are not altruistic, so wide dissemination (sales) of the products is necessary. Selection of topics for an identified "community of interest" is carried out by an editorial staff or a review and advisory board. The publication of reference books and handbooks constitutes viable, visible, and available scientific and technical data efforts of national significance. In addition to the apparent utility of these source books, the data contained in them have been refined, evaluated, and have been proved significantly sound through years of usage.

The products of technical book publishers fall into five main divisions: (1) monographs, or books which present state-of-the-art summaries and reviews, written to update the particular field; (2) reference books, or books presenting considerable data, quite detailed and specific, arranged topically and written for the materials specialist in a given science or scientific area; (3) handbooks, such as Perry's "Chemical Engineering Handbook" and Kent's "Mechanical Engineering Handbook," consisting largely of data with little descriptive text, with references to original sources less frequent than in reference books; (4) encyclopedias, which serve as guides, with very limited detail of information and data, as in the case of the "Encyclopedia of Chemical Technology;" and (5) textbooks, which are similar to monographs, but are written mainly for students, so that the style of presentation is different, and containing considerable text and discussion information, with only limited amounts of data.

Of these types, only reference books and handbooks are considered primary materials data sources. All others are designed to present ideas, theories, and concepts, rather than large amounts of data. The data contained in these are used to illustrate and emphasize the concepts and thus are limited in scope.

The fourth of seven classes of materials data sources is trade associations, which are defined as organizations establishing to serve a specific industry through individual company memberships. Liaison is maintained among the company members. Information and data are collected, collated, and organized for dissemination to participants. Joint action on various industry-wide issues is initiated and performed, sometimes by the corporate staff and sometimes by appointed committees. Trade associations with common problems often join together to obtain data of importance to the several associations. A typical example is the joint effort of the Society of Plastics Industry and the Manufacturing Chemists Association to obtain the requisite data on fire and flame resistance of "plastics for building" in order to obtain approval for their use from Building Code Boards.

The quality and quantity of materials data generated by trade associations varies, depending on the nature of the particular material and on the breadth of demand for data. Most trade associations in the materials field have been formed to promote a specific material or group of materials. Typical examples are the American Iron and Steel Institute and the American Concrete Institute. Some publish good technical journals and bulletins containing many technical data; others hold annual conferences at which technical papers are presented which are published as proceedings. All of these efforts lead to exchange and dissemination of valuable data with limitations only of circulation and of scope.

Trade associations are organized and operated to satisfy the needs of their members, who are the users of any data generated and disseminated. These user-members determine the patterns of operation and control the efforts of the associations to increase the available data and to expand the use, thereby, of the specific material. A major function of many trade associations in the materials field is determination of standards and test methods for the materials and end-products. These standards are frequently approved and distributed through appropriate

government agencies, such as the U.S. Department of Commerce. Another important operation of trade associations is the collection of statistical economic and production data on the industry. Usually, these data are considered so proprietary that no distribution is made beyond the participating members.

There are many trade associations in the United States. Each one serves a special segment of industry and of users generally associated with a specific material. Table II-C-6 lists some of the largest and most important materials trade associations in the United States.

The fifth of the seven classes of materials data sources is trade journals and magazines, which are defined as those publications designed for specialized groups of readers, largely for industrial and commercial outlets in specialized areas such as plastics. Frequently, trade journals and magazines are circulated free to the selected readers in order to obtain visibility to controlled audiences for their advertisers, which for the most part bear the costs of publication. The most valuable data published in trade journals are collected by staff solicitation. Only a fraction (estimated 20%) of total data collected by publication staffs is published, so many of the trade journals have extensive "reader service" activities to supply answers to readers' requests. Trade journals and magazines primarily utilize industry sources for their data and information, in order to provide new and up-to-date offerings to their readers; this assures a broad spectrum of reader interest ranging from technical sales through operations, management, and research. Trade journals are thus considered valuable national data efforts. Table II-C-7 lists some typical materials trade journals.

It is difficult to adequately assess the total impact of trade journals on the users of materials data. There is no user of such data who can afford to ignore the current awareness value of news items, advertisements, and reported data in trade journals. Often, additional materials data are generated by prospective users who request "free samples" for testing based on advertisements with their limited but suggestive data. In order to obtain wide reader circulation, trade journals obtain large, if not all of their, publication costs from advertisers. This usually means that printed data on materials are raw, not correlated, and sometimes biased.

TABLE II-C-6. REPRESENTATIVE MATERIALS TRADE ASSOCIATIONS IN THE UNITED STATES

Acoustical Materials Association
Aerospace Industries Association of America
Aluminum Association
Aluminum Extruder's Council
American Concrete Institute
American Die Casting Institute
American Gold Association
American Institute of Steel Construction
American Iron & Steel Institute
American Pulpwood Association
American Tin Trade Association
American Zinc Institute
Asphalt Institute
Association of Iron & Steel Engineers
Building Research Institute
China Glass & Pottery Association of America
Clay Products Research Foundation
Copper Development Association, Inc.
Cork Institute of America
Glass Container Manufacturers Institute
Gypsum Association
Industrial Diamond Association of America
Insulation Manufacturers Association
Iron & Steel Engineers Association
Iron & Steel Institute
Lead Industries Association
Leather Association
Magnesium Association
MICA Industries Association
National Concrete Contractors' Association
National Lime Association
Portland Cement Association
Refractories Institute
Rubber Manufacturers Association
Society of the Plastics Industry
Society of Wood Science and Technology
Steel Founders' Society of America
Stone Institute
Sulphur Institute
Textile Research Institute
Tin Research Institute
Tungsten Institute
United States Copper Association
Uranium Institute of America
Zinc Institute
Zirconium Association

TABLE II-C-7. TYPICAL TRADE JOURNALS IN THE MATERIALS FIELD

Actual Specifying Engineer
AISC Engineering Journal
Adhesives Age
American Glass Review
American Metal Market
America's Textile Reporter
Architectural & Engineering News
Biopolymers
Blast Furnace & Steel Plant
Brick & Clay Record
Building Science
Carbon
Ceramic Age
Ceramic Data Book
Ceramic Industry
Ceramics Monthly
Chemical Engineering
Concrete Construction
Construction Methods & Equipment
Cotton Trade Journal
Electronic Products
Engineering Alloys Digest, Inc.
Glass Digest
Insulation
Iron Age
Lubrication
Materials Engineering
Materials Handling Engineering
Metal Working
Modern Materials Handling
Modern Plastics
Modern Textiles Magazine
Plant Engineering
Plastics Design & Processing
Plastics Technology
Plastics World
Product Engineering
Progressive Architecture
Public Works
Reinforced Plastics
Rubber Age
Rubber World
Steel
Textile Bulletin
Textile Organon
Textile World
Western Plastics

The sixth of the seven classes of materials data sources is professional societies (or institutes) which are defined as organizations oriented towards a community of interest based upon common scientific or technical desires. TableII-C-8 lists primary materials-oriented societies. Some societies are organized to serve a common discipline, e.g., the American Chemical Society. Such societies are composed frequently of Divisions wherein the members of the specific discipline display their common interest in a given group of materials, e.g., the Polymer Division of the American Chemical Society. Other societies are organized to support a community with interest in a specific group of materials, e.g., the Society of Plastics Engineers; in these, members are bonded together by their common interest in the materials and are drawn from many separate disciplines. Membership requirements of most professional societies are stringent and emphasize formal education, as well as training and experience. The professional societies, by historical precedent, have accepted major responsibilities for dissemination of data and information. Formally, these efforts are directed towards conferences and publications. At conferences, scientific and technical papers are presented, new research efforts as well as applied technologies are discussed, and broad interchange of information and data is encouraged. Preprints of individual presentations and conference proceedings are frequently made available to each attendee. Information and data are thus made available months, if not years, before their publication in journals.

Professional societies have since their beginnings been the major sources of scientific and technical journals. This type of publishing effort originated with the early European societies and was adopted in the United States. The larger professional societies publish many journals that carry data ranging from highly sophisticated and theoretical researches to the more applied and practical efforts. The American Chemical Society publishes an excellent research journal, Journal of the American Chemical Society; it also publishes more practical journals, such as Industrial and Engineering Chemistry. The review of articles before publication is usually quite thorough and is done frequently by members of the professional society who are quite familiar with the subject of the article. This review procedure assures the necessary refinement and evaluation of data under strong editorial policies. Costs of publication of professional society journals are borne jointly by subscriptions, society contributions, and advertising revenues; however, with the constantly increasing costs of publication, the costs have been disproportionately thrown onto advertisers and onto government subsidies.

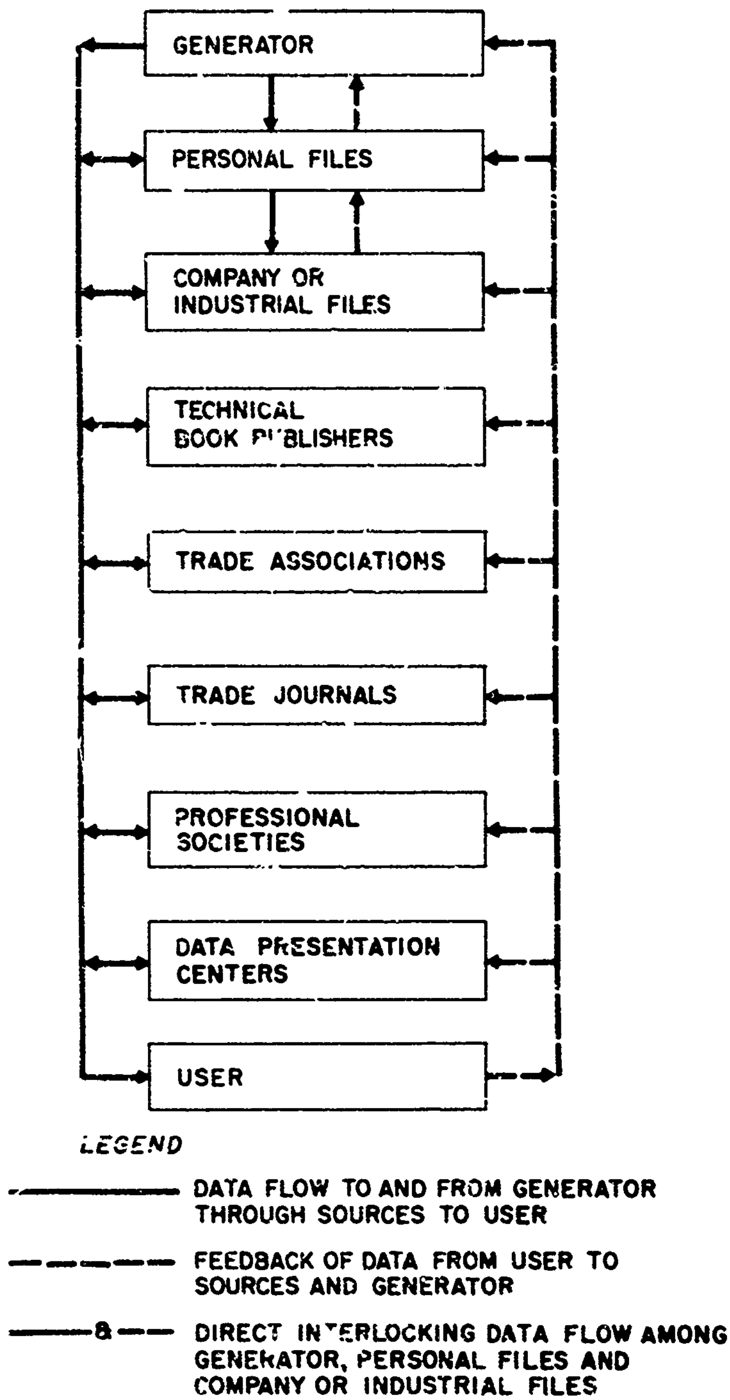


Figure II-C-2 Simplified Materials Data Flow

TABLE II-C-8. PRINCIPAL PROFESSIONAL SOCIETIES IN THE MATERIALS FIELD

Acoustical Society of America
American Ceramic Society (ACS)
American Chemical Society
American Foundrymen's Society
American Institute of Aeronautics and Astronautics
American Institute of Architects
American Institute of Chemical Engineers
American Institute of Mining, Metallurgical and Petroleum Engineers
American Paper Institute
American Petroleum Institute
American Society of Civil Engineers
American Society of Heating, Refrigerating and Air Conditioning Engineers
American Society of Lubrication Engineers
American Society of Mechanical Engineers
American Society for Metals
American Society for Testing and Materials
American Society of Tool and Manufacturing Engineers
American Welding Society
Association of Iron and Steel Engineers
Brass And Bronze Ingot Institute
Construction Specifications Institute
Data Processing Management Association
Electrochemical Society
Gray and Ductile Iron Founders' Society
Institute of Electrical and Electronics Engineers
Institute of Environmental Sciences
Instrument Society of America
Manufacturing Chemists' Association
National Association of Corrosion Engineers
Society of Automotive Engineers
Society of Plastics Engineers
Technical Association of the Pulp and Paper Industry
United States of America Standards Institute

The seventh and last of the seven classes of materials data sources is the formalized data efforts, such as data centers. The collected data may be provided from such data efforts in their original form and/or refined form on request. More sophisticated facilities evaluate and otherwise refine the data for dissemination. Presentation for dissemination varies from duplication of stored data references to highly organized and sophisticated handbooks. In between these two extremes, there are variations of refinement and evaluative techniques that result in degrees of reliability of the disseminated data. The value of any data center in the materials field depends upon the needs and sophistication of the user population.

Various stages of evaluation are recognizable in data presentation centers. Collection of data is an inherent part of the operation of most data facilities. The collected data may be either evaluated prior to storage, or may be stored in its raw, unevaluated form. Storage of non-evaluated data is more expensive and usually of less value on retrieval. Any storage system, to be operable, must have a built-in retrieval capacity. The retrieved data may be presented and disseminated in various ways, such as edge-punched cards for manual sorting, machine-sortable cards, magnetic or photographic systems. Materials data centers may be classified into two general categories:

- One grouping is concerned with a specific type of materials and its properties. Typical of a material-oriented center is the Tin Research Institute, Inc., which is supported primarily to promote the use of tin and tin-containing materials. Many such data centers are industry-supported to promote usage of the particular material and thus increase profitable sales volume.

- The second category of materials data facilities includes those which collect and coordinate data concerned with specific properties. In this group, there is relatively little direct industry support. The facilities are usually wholly supported by government funding, either "in-house" or under contract. Typical of this type of activity are such centers as the Air Force Machinability Data Center, the Cryogenic Data Center, and the Engineering Materials and Process Information Service (EMPIS).

A limited survey conducted among professional personnel who are direct users of materials data indicated that only one out of three was aware of formal data efforts. This suggests that visibility of data centers needs marked improvement. Only one out of six had tried to use a data center, and use was discouraging because of lack of accessibility (availability) to the desired data. This suggests that a better system of accessibility be developed in order to broaden the usage pattern. A list of formal data efforts in the materials field is contained in Table II-C-9. These are typical of efforts relating to a specific material or material family, as well as those relating to property categories. Complete details on several of these are given in Part C of this volume.

There are many relationships between sources and generators of materials data. Many users are per se generators of materials data which may be confined to their own personal files (perhaps unwritten) or, through a feed-back mechanism, may become visible and available in other source data accumulations. Therefore, there is no simple path of materials data flow starting with the generator and ending with the ultimate user. Many intermediaries may intervene between generator and user, or the user generates his own data. Therefore, any attempt to completely describe data flow from generator to user is most difficult. Figure II-C-2 is a simplified materials data flow model that shows the major sources of data, as well as the types of data usually associated with each source; it shows materials data flow from the generator to the various users and the feed-back of data to the sources.

4. Problems in Materials Data Management

One of the major problems in management of materials data is the sheer volume of such data. Part of this volume arises from the increasing knowledge of old materials; another part is due to the increasing development of new materials, as well as the demands for improved properties. This increase in varieties of engineering materials was emphasized in 1960 by van Vlack (Lawrence H. van Vlack, "The Two Major Trends in Materials Education," Materials in Design Engineering, pp. 151-155, September 1960). Figure II-C-3 shows this increase graphically. Using the number of varieties of materials in 1900 as "X", it is predicted that the number in 1975 will be "10,000 X."

TABLE II-C-9. A LIST OF DATA PRESENTATION CENTERS FOR MATERIALS

Aerospace Materials Information Center
Air Force Machinability Data Center
Ceramics & Graphite Information Center
Cobalt Information Center
Copper Development Association Technical Data Center
Cryogenic Data Center
Defense Metals Information Center
Electronic Component Reliability Center
Electronic Properties Information Center (EPIC)
Engineering Materials & Process Information Service (EMPIS)
Fused Salts Information Center
Infrared Spectral Data Center
Liquid Metals Information Center
Mechanical Properties Data Center
Metal Plating and Coating Information Center
Non-Destructive Testing Information Center
Plastics Technical Evaluation (PLASTEC) Center
Radiation Effects Information Center
Rare-Earth Information Center (RIC)
Research Materials Information Center (RMIC)
R and D Technical Information Center
Superconductive Materials Data Center
Thermodynamic Properties of Metals and Alloys Center
Thermophysical Properties Research Center
Transducer Information Center (TIC)
Tungsten Institute

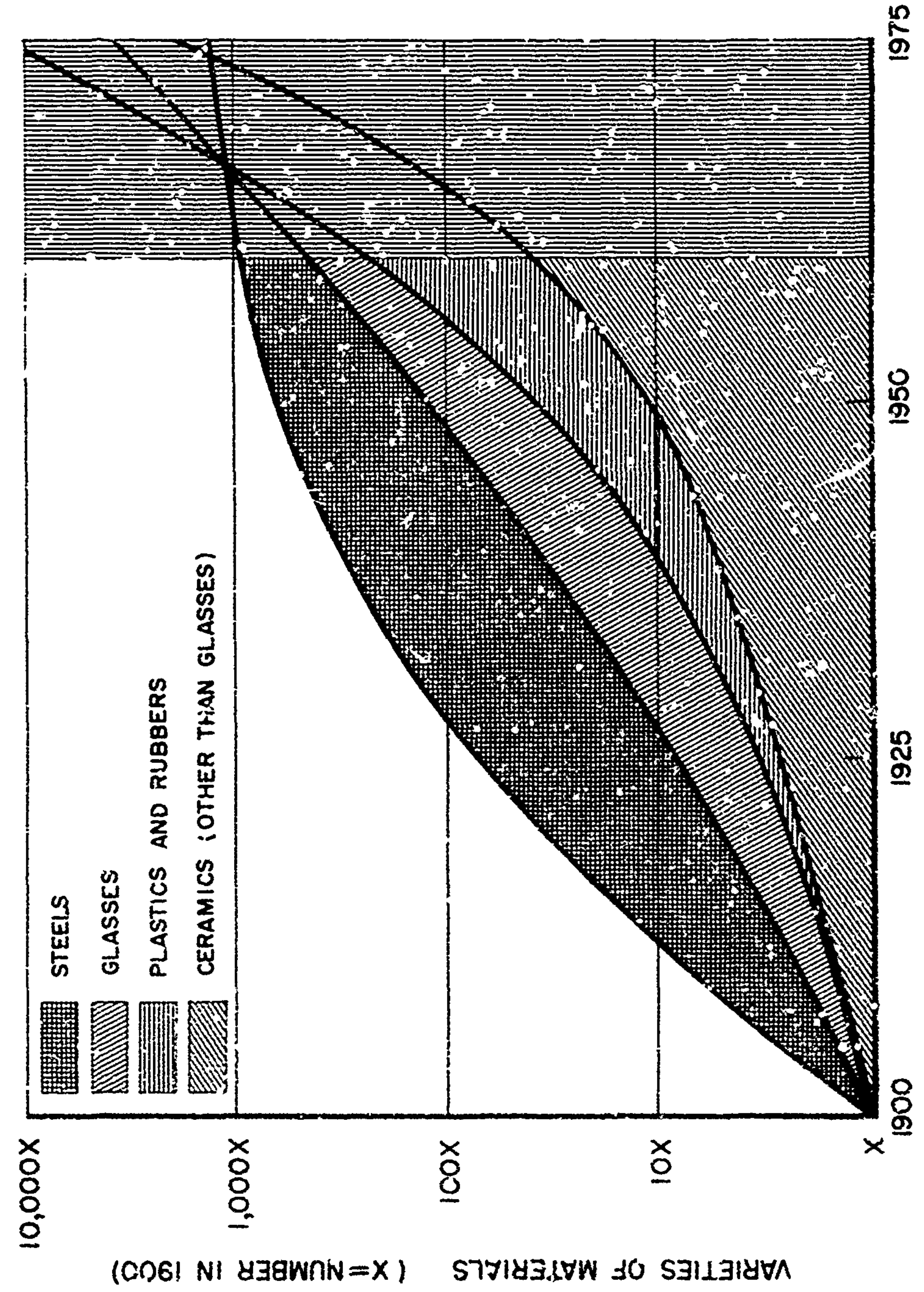


Figure II-C-3 Increase in Varieties of Engineering Materials Since 1900

N. E. Promisel ("The Role of Materials and MAB in Technology," Materials as a Common Denominator in Engineering Achievement, pp. 1-13, National Research Council Bulletin, 1967) explains this phenomenon of the increasing rate of development of new materials:

> "Since everything is made from materials, we are dealing this afternoon with an infinite subject. And because it is all around us, materials are taken for granted, as a quite pedestrian topic -- until suddenly the "March Of Progress" in another field is stopped, the "Frontiers Of Achievement" somewhere else are no longer pushed forward -- because just as suddenly some designer realizes that there is not suitable or efficient material from which to built his new gadget, his new device, his new supersonic transport, his new high speed transportation system, his new artificial kidney, his new submarine for underwater exploration, his new high power rocket for space flight, or what have you. Then one of two things could happen: he could give up temporarily and wait five to ten years for a new or improved material; or he could build his gadget with sacrifices in performance, or less efficiently, and end up with something that would be obsolescent by the time it is put into service. Usually, it's a crisis, and now the materials requirements are no longer mundane and pedestrian but exotic and sexy, and crash programs, always expensive, are initiated to eliminate the crisis. All this happens because 'we haven't planned ahead,' because the designer and materials engineer have not properly appreciated and dealt with their 'interface,' and because science and engineering have not interacted promptly or adequately. Obviously, there is some room for improvement."

The increasing number of varieties of materials, the expanding new requirements for modern exotic applications lead to production of an enormous volume and increased complexity of materials data. This produces a problem that is not soluble by a centralized storage and dissemination system. It is easily visualized that a centralized system would require a bureaucracy greater in numbers than that of the Internal Revenue Service and one with a much higher degree of training and competence.

A major factor is the interplay among materials and the dependence of each given material upon the others, further complicating the overall problem of management of materials data. This interdependence is well described by consideration of reinforced concrete. This composite material gives properties that are synergistic accumulations of the properties of the steel reinforcement and of the concrete matrix.

Promisel (loc. cit.) discusses the efforts of the Materials Advisory Board to characterize materials. His statement emphasizes the overall complexity of materials data:

> "Since there are no materials completely devoid of contaminants and no perfect crystals, unless a material is adequately identified with respect to its composition and structure, interpretations of measured properties must be viewed with severe reservations. This may seem obvious, particularly to those of us in the materials field, but too much of the research done today, both basic and applied, is performed on materials taken for granted rather than adequately characterized, thus seriously degrading the usefulness of the results. Catastrophic failures have occurred in service because of inadequately characterized materials. The flow diagram for materials is illustrated in Figure 4. On the left are the starting ingredients and procedures which determine the materials composition and structure, including

> defects. The composition and structure then determine the properties and uses. The properties do not characterize the material; the converse is true. Thus, we are led to the definition that true and ultimate characterization describes those features of the composition and structure (including defects) of a material that are significant for a particular preparation, study of properties, or use, and suffice for the reproduction of the material. True characterization is the cornerstone of material science and the committee studying this concluded that this was a vital message that had to be repeatedly impressed on all scientists and engineers, in all fields. Particularly, solid state physicists seem often to have ignored this. Those of you who have influence over research and those of you who review research reports for publication, especially those representing professional societies, would render an important and needed service if you would keep this fundamental requirement in mind, and, where appropriate, reject submitted papers describing work on inadequately characterized materials."

Another problem is that any experienced user of materials data, when confronted with new data on an old material or with data on a new material, raises questions concerning the quality of the data: What is their source? How were the data taken, i.e., what were the test conditions? What was the composition of the material? What was the process of manufacturing? How exacting were the quality control standards? Did the generator of the data possess the requisite skills and experience? In other words, the user, whether researcher or engineer, must know how reliable are the data and what confidence can be placed on the data. This lack of confidence in some materials data by the potential users leads to two discrete problems:

- Users often discount the value and reliability of data on materials. This feeling of insecurity

about the materials data forces designers to use higher-than-necessary safety factors that give heavier structural elements. While the heavier-than-needed structures are of little harm on earth-bound designs, they become serious handicaps in dealing with aerospace or hydrospace structures; and

- Users of materials data are known to doubt data acquired from competitive facilities, and they frequently doubt data obtained from other divisions of their own facility. This leads to an extensive duplication of materials testing and experimentation and causes a redundancy of data generation.

In this context, the lack of confidence in reliability of materials data is most noticeable in design with new materials. This is particularly true if the new materials are composites. The data on the older, more conventional materials are considered much more reliable by users.

Another problem is the need for education in the use of materials and in the interpretation of materials data essential to proper and safe design of structures and artifacts. There is an increasing awareness of the need for expanding formal training and education in materials, both research and engineering. This need is manifest by the rapidly increasing numbers of collegiate level courses in polymeric materials. Winding and Brodsky ("SPE Education Committee Survey of Polymer Courses," C.C. Winding and P.H. Brodsky, SPE Journal, 24, No.1, 31, 1968) list 106 universities that were teaching one or more courses in polymeric materials in 1967, compared to only 37 in 1950. Of this number, 23 offered 20 or more separate courses in this materials area. Formal education in the use of materials and materials data requires thinking in light of training and experience. The safety and welfare of peoples, as well as security of nations, demand judgments in choice of materials based upon property data. Helmreich ("Some Thoughts on Education," Jonathan Helmreich, Allegheny College Bulletin, Winter 1967-68) presents some excellent concepts on materials education and data usage:

> "Education does not mean you know all the answers - indeed it should convince you of

just the opposite - but the educated man will know how to go about finding the answers. Moreover, the educated man realizes that finding data is not necessarily equivalent to finding the answer. It often helps a great deal, but still something more is needed. Analyses of various treaties and diplomatic incidents may show how it was possible that World War I came about, but not <u>why</u>. Socio-economic studies of Watts, Hough, and Harlem may bring forward impressively frightening statistics, but they do not tell why riots have or have not occurred, and they certainly do not inform us as to the proper course to avoid trouble - or even whether it might perhaps be better that riots occur than that seething hostilities be further repressed.

"If anything is to be done with data, if all the information that is being crowded onto computers is to have meaning, then questions must be asked. And the asking of questions - precise, pertinent questions - is the supreme mark of the educated man. Moreover, it is the only way of becoming educated. There are many tricks to becoming trained, but it is only the rocky and lonely path of questioning that will lead to an educated outlook on life. The slips and falls that are taken even on its first turning are such as to discourage faint hearts. I should like to think it will not discourage you. But it is baffling to be confronted by some boulder of a problem. Many are scared by its sight; others will make only a feeble attempt to scale it and then turn aside saying it is not worth the effort. Only a few truly strive to force their way past the obstacle. The path is lonely, because only one man can travel it at a time. Others have gone before, and as teachers, they will try to show you the handholds that will help you along. Do not fail to make use of them - nothing is more discouraging than a class with no questions, or an advisee who fails to come by to talk with his adviser. Yet

> essentially you must make your way yourself. It is you that must face up to the agonizing realization that you really don't understand yourself, your neighbor, the problem you are confronting."

The problem in data management in materials becomes one of not formal education, but of ability to think. This reduces to a simple minimum of inquiry: Can data management, no matter how sophisticated, help to bridge the gap between the educated, but inexperienced generator of materials data and the similarly educated, but non-thinking user?

Another major problem in management of materials data is the cost; i.e., dollars per unit of data; this is closely related to the volume of data, as well as the complexity of operations. Historically, the major cost of publication of materials data was borne by members of professional societies, subscribers to the various publications. As publishing costs increased, dues and subscriptions ceased to supply the additional revenue required for society publications. Professional societies were forced to turn to increased advertising or to governmental subsidies through contract studies. Survival of traditional sources of materials data is essential because, in most cases, the published data are highly evaluated and reviewed by peers in the area of direct interest. One answer to the problem is data dissemination by trade associations, supported by industrial members so that the cost is distributed on a broad basis. However, this policy of support means that the distribution base is limited and the materials data may suffer distortion. Another answer is a major shift to trade journals, wholly supported by advertising revenue. Thus, published materials data are reported to be independent of bias due to advertising funds, but the quantity and accessibility of such data that can be published is controlled by the total revenue available from advertisers. This merely means that no competent businessman will publish data on any material at a loss. Furthermore, due to costs, the individual user of materials data cannot afford to subscribe and pay the publication costs personally. In many cases, university and small facility libraries have to forego subscriptions because of limited funds. For example, a plethora of publishing efforts has arisen in recent years, designed to aid the flow of data, including materials data; these efforts are available at a cost prohibitive to an individual. Typical is CCH's Clean Air News, which promises 52 issues for $48.00 per year; this is published by Commerce Clearing House, Inc. of Chicago, Illinois.

Related to the problem of data cost is the proprietary aspect of materials data generated by two major categories of industry: (1) Producers and suppliers of materials; and (2) Manufacturers and fabricators of end-product items. As mentioned earlier, data generated by suppliers of materials are usually visible and available; suppliers willingly distribute these data in order to further acceptance of their materials for construction, manufacturing, and fabrication. On the other hand, manufacturers and fabricators of end-product items hoard data to preserve a perhaps false, competitive advantage. Composition and processing data for materials, as well as fabrication procedures and techniques, are proprietary, and these data are neither visible nor available. The company-funded data-producing efforts are proprietary; results of materials data are also proprietary. Employees of all large companies in technical areas are required to sign an agreement that protects the secrecy and proprietary nature of any data produced; this agreement normally remains valid for one year after termination of employment. The importance of the proprietary data to a total sum of materials data is unknown, as these data are not available for survey, but it is feasible that these constitute the bulk of existing materials data. While it is difficult to imagine how this latter problem might be solved, it is essential that the formerly stated problems receive further study for their resolution.

D. Chemistry and Chemical Engineering

1. Introduction

Chemistry and chemical engineering is defined as the science and technology that deal with the composition, state, and properties of substances; and the phenomena and processes by which they are transformed. It includes the subfields of analytical, physical, inorganic and organic chemistry and chemical engineering, all of which are interrelated. Chemical engineering is defined as the technology and industrial implementation of chemical transofrmation phenomena and processes.

As used here, chemical engineering does not include: mechanical design of processing equipment, the management aspects of processing, the end use of chemical products, the technical functions which fall within the scope of nuclear reactor engineering, petroleum and mineral exploration and exploitation, or materials engineering. Materials engineering is limited to the study of the macrostructural properties of solids, particularly those which pertain to product design and fabrication. The materials industry is thus defined as the industry sector which is concerned with manufacture of solid materials (plastics, metals, etc.) for product manufacture. In contrast, chemistry and chemical engineering are concerned with both the macrostructural and microstructural properties of solids, liquids, and gases.

The chemical process industry includes the processing sectors of the following industries: food; textile; paper and pulp; chemical products (drugs, industrial chemicals); petroleum; rubber and plastics; stone, clay and glass; and extractive metallurgy. Processing aspects of food manufacture, chemical fertilizer, and pesticide production are included here: food, feeds, fertilizer, and pesticide formulation and application aspects are covered in the "Agriculture and Food Technology" section.

When the more fundamental terms are employed in expressing such distinctions, the field of chemistry and chemical engineering reveals itself as a concern with substances. In Table II-D-1, which contrasts

TABLE II-D-1. THE DOMAIN OF CHEMISTRY AND CHEMICAL ENGINEERING:*
THE WORKING DEFINITION FOR DATA DISCUSSED IN THIS SECTION

CHEMISTRY	(versus PHYSICS:)
Properties of elements and compounds (molecular weight, refractive index, boiling point)	...v. Properties of matter in general (mass, inertia, etc.)
Energy changes from alteration of composition and/or state (heat of vaporization, spectral emission, heat of formation, radioactivity)	...v. Forces related to matter in general (gravity, centrifugal force, etc.)
Composition, transformation and structure of substances, mixtures, molecules, ions, atoms, and other elementary particles.	... (No analog)
Interactions of energy with matter (induced radioactivity, absorption, scattering, etc.)	...v. Electromagnetic radiation per se.
	(versus MATERIALS SCIENCES AND ENGINEERING:)
Basic macro- and micro-structural properties of substances.	...v. Macrostructural properties of solid-phase materials, particularly those of economic interest
	(versus BIOLOGY:)
Structure of biological substances and effects of specific substances on biological substances and systems.	...v. Non-specific substances, effects not characterized by specific substances
	(versus EARTH SCIENCES:)
Substance-specified compositions and phenomena.	..v. Physical phenomena
CHEMICAL ENGINEERING	(versus OTHER ENGINEERING AND TECHNOLOGY:)
Physical aspects of heat and mass transfer	...v. Formulation, industrial processing, and commercial application of mixtures of substances
Technological concepts for accomplishing physical and composition change of substances, including size-scale influences.	Management aspects of processing and commercialization
Industrial processing, including chemical fertilizers and pesticides, and commercial application of substance change.	Technological concepts not contingent on substance identification or interaction (scale-up modeling, etc.)

*Developed from "Directions for Abstractors", Chemical Abstracts Services, 1967, with modifications that emphasize scope distinctions for other science-technology sections of this report.

chemistry and chemical engineering with several other fields, the most frequent point of distinction is the identification of substance, the change of substance, the property of substance.

One may perhaps be near the distinguishing essence of things "chemical" with the concept of change of substance. Here the scientist has left the domain of the stable, the inert, the permanent, that can be measured once, counted once, and thereafter manipulated as a stable entity where, when and how he pleases. When water changes from the liquid to the vapor phase, most of the consequences can be dealt with only through data specifically concerning that phase change of that substance.

If the substance is ethyl alcohol, substance-specific data for ethanol are needed; and if the liquid-solid phase relationship is called for, substance-specific data for the property change are needed. If a mixture is involved, such as ethanol-water, its liquid-solid phase relationship cannot be established from a simple arithmetic manipulation of those properties pertinent to each substance; the substances have interacted with each other, and the result is a property pattern unique to that two-constituent system.

When the scientist's preferred realm of pure substances is unattainable because of industrial economics, or even of limitations in purification techniques, slightly impure substances usually behave much like pure substances: however, one may discover that some properties of slightly impure materials may differ radically from those of the pure substance -- in other words, the property characteristics are those of an entirely new multi-constituent system. One dramatic example, semi-conductor materials, is the basis of a significant industrial art that revolves about the commercially significant consequences of creating electronically interesting substances through meticulous control of minute proportions of additives to a base substance.

When the chemical engineer's processing scale extends beyond the test-tube, the reaction heat of the constituents... coupled with the geometry of the reaction vessels...the thermal diffusivity of the system comprising the reaction batch and the container --- and the differing rates of substance change for different substances in many compositions at different temperatures--- (and so on) --- may constitute

a multi-parametric system that is not even worth characterizing as an explicit assemblage of basic "chemical" factors, even if it were found possible analytically. Instead, the process engineer's primary measurements and documentary records of substance change may be expressed as the system-envelope response to a changed input parameter, such as the flow rate of a process reactant (or even of the coolant to one heat exchanger).

In summary, the field of chemistry is that of a phenomenon significantly associated with virtually all the technical arts and natural sciences. The basic conceptual structure of the science was well-established over two centuries ago. It has been one of the technical fields most populated with scientific manpower, and one in which the effects of a strong and long-standing research tradition can be seen in a sophisticated and highly versatile technological competency. This competency is a pervasive resource: it serves all the technical fields as an important tool, as well as forming the cornerstone of the industrial sector termed the "chemical process industries". As a basic science, chemistry flourishes. Chemical research activity shows no evidence of exhausting soon the potentials for further knowledge of the phenomenon and its possible application.

Some Measures of the Field of Chemistry and Chemical Engineering - Because chemical personnel and facilities will be found in virtually all major technical institutions, available national manpower statistics probably provide one of the most useful general measures of the proportion of current science-technology activity that is chemically oriented. Some of the more significant ratios follow:

In 1966, chemists comprised 28% of an estimated scientist population of 500,000, which was over twice the percentage represented by the next largest scientific field (biological sciences). Chemical engineers comprised 10% of an estimated engineering population of about 600,000. These chemical professionals were distributed as follows

among the three key employer groups:

	Percent Employed In			
	Industry	Government	Academic	Other
Chemists	56	6	22	16
(All Scientists)	(34)	(10)	(36)	(20)
Chemical Engineers	92	2	4	2
(All Engineers)	(71)	(15)	(6)	(8)

R&D scientists and engineers in the chemical process industries, as represented by the "Chemicals and related products" of the census of manufactures, comprised 11% of the R&D professionals in all manufacturing. Since total manpower in the chemical sector was only 4 percent of the total for all manufacturing, it is evident that chemically-oriented industrial activity operates at a generally high level of technical sophistication.

Bibliographic statistics provide perhaps as accepted a measure as any for the rate at which new chemical knowledge is being generated. An estimated total of 200,000 papers and reports and 100,000 patents containing new chemical information appeared in 1966. The over-all annual growth rate for such items is 9 percent, compounded, which projects to a level of almost 400,000 chemical documentation items generated annually by 1970. This input of new knowledge adds to a current accumulated total (as measured by the combined coverage of Chemisches Zentralblatt and Chemical Abstracts) of approximately 4.5 million items. Chemically significant information is estimated to appear in 30 percent of the world's technical journals, and in 15 percent of all currently issuing scientific and technical papers. Information specifically identified by substance (i.e., the molecule, structure, reaction, etc.) is estimated to comprise 85% of this output.

The language of chemistry benefits from a strong and extensively articulated vocabulary. The most elementary and most rigorous portion of this language -- the properties of substances in defined

environments -- is the terminology in which chemical data are expressed. Approximately four million different chemical identities are now known, and approximately 100,000 additional compounds are discovered or created annually. Various authorities have identified between 1,000 and 2,000 discrete properties as being of scientific and technological interest. Since most of these substance-property pairs can be identified readily in relation to the environmental parameters (e.g., temperature, pressure, radiation, etc.) that are acting individually or in combination to affect the substance at the same time, an extensive framework exists for recording chemical experience in terms that are readily described and manipulated by technically literate individuals.

Roles Chemistry Plays in Other Scientific and Technological Fields - To the extent that the composition of substances is significant to a field of science or technology, chemical concepts, the chemical data that characterize the properties of specific compositions, and the compositions themselves are significant. The geophysics of the atmosphere, which at first glance might appear a field having little connection with chemistry, provides an informative illustration:

- To estimate the input of water vapor to the Earth's atmosphere requires knowledge of the vapor pressure of the liquid aqueous phases over the temperature and salinity ranges encountered.
- The radiation absorption and emission properties of water and the atmosphere, including such constituents as carbon dioxide, are used in establishing the thermal flows, phase and energy conversions, and balances that help explain weather patterns over wide geographic reaches.
- The concentrations, composition, form, and distribution of particulates in the atmosphere help explain the role of condensation nuclei in precipitation.
- The density-temperature relationships help explain the vertical circulation of sea water in the oceanic basins

Other examples could be given.

The fact that this familiar basic chemical knowledge tends to lie beneath the gross phenomena of geophysics should not obscure its importance as a bridge between the descriptive and analytical levels of geophysical knowledge. Without the rigor of the "hard-science" disciplines such as chemistry, the description of a cloud would hardly be as precise as that of an angel.

Applicable chemical information also is one of the important factors that advance technologies to more effective levels. The molecular explication of Vitamin A, followed by the development of an economic manufacturing procedure involving chemical syntheses, was a welcome as well as profitable successor to the more empirically-grounded fish-liver-oil process. Transistor compositions are an instructive sermon on what can result when an electronic process, long recognized but until recently not understood in the mineral galena, and understood but inefficiently accomplished in the electron tube, finally can be accomodated with great effectiveness through "designed molecules".

Yet another aspect of chemical knowledge should be pointed out in this discussion of the roles it plays in other scientific fields. Certain interacting complexes of chemical substances and properties are so expressive of the human interest in the phenomena involved that chemistry has lent its name to them. In the chemical literature, the data expressing these phenomena are given not in basic chemical units but through such "property" terms as Biochemical Oxygen Demand, fuel specific impulse, Octane Number, citrate-soluble phosphate, and the like. In Table II-D-2, the summary lists illustrative of chemically-oriented technical activity in the aerospace and agriculture fields suggest how widely the chemically specialized dialects range at the technological level.

At the industrial level, many of the products of the chemical process industries are ingredients or components of other technologies, rather than end-products. Examples are particularly recognizable in the fuels, detergents, lubricants, and protective coatings that associate with many industrial processes and consumer products. Chemical technology has generally proved capable of responding well to the demands other technologies have placed on it for chemical products possessing specified properties.

TABLE II-D-2

ILLUSTRATIVE CHEMICAL AND CHEMICAL ENGINEERING SUB-FIELDS ASSOCIATED WITH RESEARCH, TECHNOLOGY AND OPERATIONS IN THE FIELDS OF AEROSPACE AND AGRICULTURE

AGRICULTURE	AEROSPACE
Basic Phenomena	
Biochemistry of plant and animal metabolism	Reactions of highly energized compositions or energy-dense systems
Constituents of biological substances	Physical and reaction properties at low temperatures and pressures
Nutrition chemistry	Composition and properties of the atmosphere
Constituents of soils	Equilibrium and phase chemistry
Nutrient diffusion and reaction in soils	
Technological Elements and Concepts	
Fertilizer chemistry and application	Fuel and propellant chemistry and applications
Pesticide chemistry and application	Combustion chemistry
	Dynamic environments and influences
Technological Operating Arts and Concepts	
Fertilizer manufacture	Propulsion system design and manufacture
Pesticide manufacture	Fuels processing
Pesticide chemicals manufacture	Propellant processing and fabrication
Fertilizer and pesticide testing	

A significant and growing fraction of commercial chemical production consists of relatively pure compositions, rather than mixtures. The plastics industry, for example, can select from a wide inventory of chemical resins and plasticizers to create materials with the desired characteristics. Backing these products up, the companies producing them can apply the insights of chemical science rather directly through their manufacturing arts, to create additional basic products for the plastics industry almost on demand. The chemical industry thus possesses a science-based technological capability which is rather unusual at the production level. This capability may explain the ready acceptance and use of chemicals and chemical methods in most modern technologies.

The potency of chemistry as a tool for the technologist is explained in large measure by the power it provides him to identify, select, or actually create compositions whose physical or reactive properties are of practical interest to him. Figure II-D-1 suggests how, through chemical steps that need only be small individual increments, major technological advances can ultimately be achieved. Most of chemistry's contributions to technology come through small simple steps. However, the masterstrokes, such as that underlying the intra-molecular complexity of transistor materials, become increasingly attainable as chemical knowledge proliferates.

2. Chemical Data Characteristics

The highly structured character of the chemical discipline provides rigorous ultimate standards for the expression of substance-property attributes. Most of them constitute a challenge to the chemist's arts of purification and measurement. Measurement limitations, in fact, may represent the practical limit of the chemist's knowledge of how pure his test sample actually is, in addition to limiting his capacity to measure its properties.

The measurement arts, as well as applications criteria, thus combine to define the realistic standards for chemical data worth conserving for uses beyond the need that supported the original

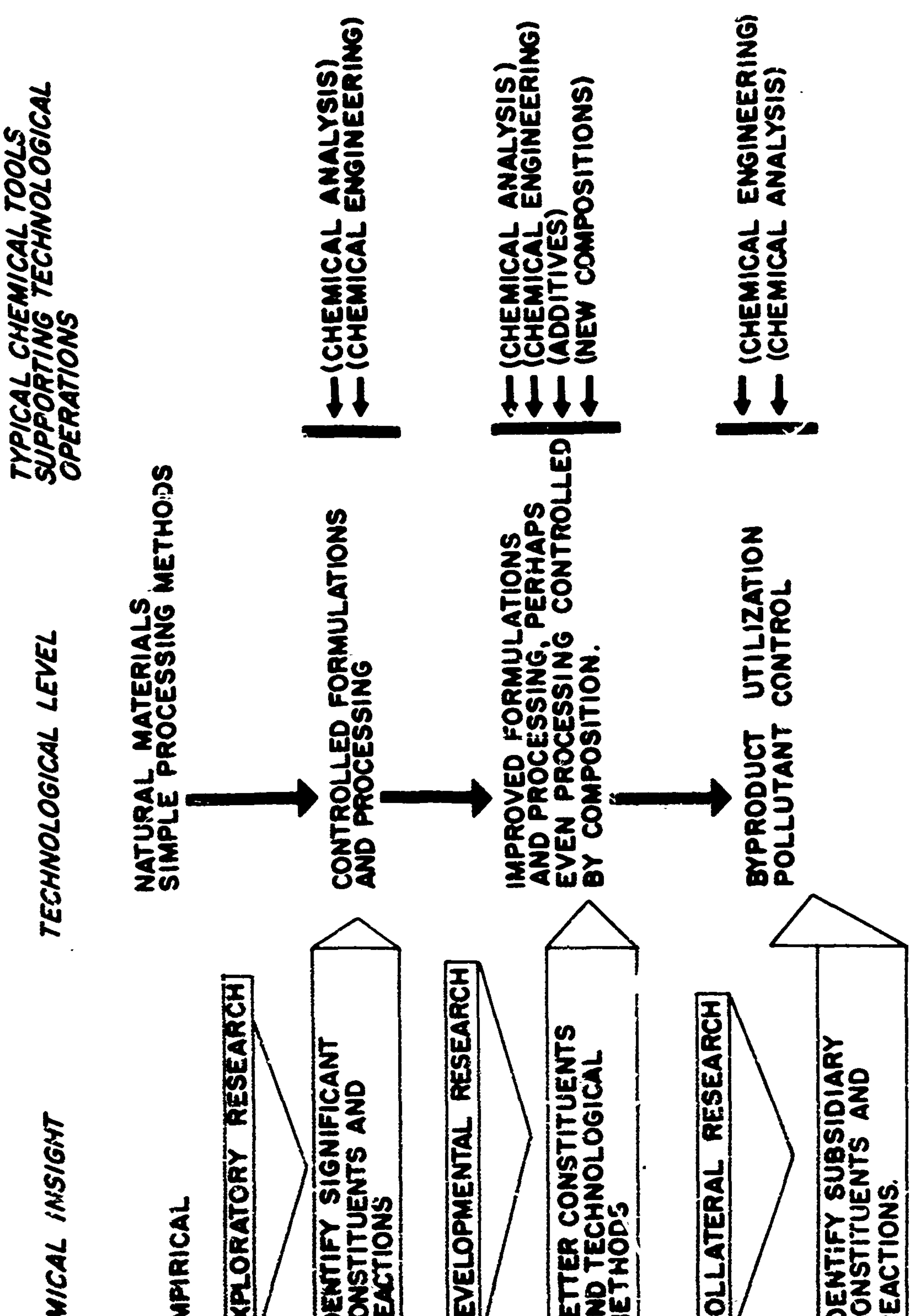


Figure II-D-1 How Chemistry May Contribute to the Advancement of a Technology

measurement. Data conserved for scientific activity may become obsolete as later data are generated that utilize samples that are more pure and that employ more accurate and precise measurements.

However, in many scientific regimes, the requirements of a study will tolerate relatively "obsolete" chemical data without compromising the observations and measurements that are critical to the investigation. This tolerance threshold allows the most regorous and knowledgeable experimentalist to utilize an old text or handbook for much of his substance-property reference data. The document is still scientifically obsolete, however, in the sense that there will be users, for each type of information in the document, whose scientific goals call for the latest and best data.

To meet the spectrum of technological needs, criteria for chemical data quality range upwards to -- and in some instances beyond -- the highest limits achievable within the present arts of sample purification and property measurement. For example, petroleum refining technology is dominated by arts associated with the chemical reactions and separation processing of mixtures predominantly composed of straight-chain hydrocarbons. Decades of basic scientific work at multi-million dollar annual levels have been invested in upgrading the quality of data describing petroleum constituents (an endeavor that has also produced great benefit for the scientific data measurement art in general). However, this body of advanced data still lacks the precision that refinery designers could use today to save millions through better-balanced designs. Another illustration can be found in the relatively recent field of rocket fuels. Here, the thermochemical laws provided a strong theoretical framework for using a high-quality chemical data base to search via calculations for better propellant combinations before embarking on the time-consuming, hazardous, and expensive route through the laboratory, pilot plant, and rocket firing bay.

Upon noting such examples, and the advances they have made to already high-grade chemical data efforts, it is tempting to speculate that technological need, rather than scientific goals, may actually power most of the actual advance in the chemical data art. It should be noted that the emergence of "high technology" is essentially the growing expression of a science-oriented style of technological in-

novation. This style is based upon (or at least requires) a strong interconnecting structure of scientific theory, which can be readily utilized through a good stock of measured or reliably calculable data. On examination, a large fraction of the data actually employed in high-technology methods appears to be substance-property, or chemical, data. It thus appears probable that the basic technological need for chemical data of high quality is not only a major need now in the more advanced present technologies, but may prove the forerunner of a general, science-oriented technological style a few decades hence.

Chemical data employed in technology range from the rigorously defined basic data required in such "paper studies" to crude and empirical measurements. The good technologist does not invest in accurate measurements for their own sake, and a great deal of the measurements in industrial chemical processing do not require high precision. The tolerances in composition and property for industrial raw materials, intermediates, and products often, of course, are very broad by comparison with chemical data required for scientific investigation or industrial research.

<u>Significant Consequences from Chemical Analysis Activity</u>. The impure materials associated with technological activity also impose the practical requirement for chemical analysis, in order to identify the kinds and amounts of significant substances in process materials. The requirement for such analyses also accounts for a large and highly significant endeavor contributing to advances in the chemical data art. The art of chemical analysis, on close examination, is the use of a property or set of properties of a substance to identify or measure its presence in a sample. Before World War II, the chemical analysis art largely involved series of laboratory manipulations. The analyst physically separated the desired composition until he secured a relatively pure compound. He could then weigh it, or react it with a reagent, permitting him to calculate its relative abundance in the sample. The large volumes of data thus generated on solubilities, reaction equilibria, and separation methods in the course of analytical research have contributed importantly to many of the industrial process methods subsequently developed by chemical engineers. The "wet analysis" art generated data that quite clearly flowed into the innovative sectors of industrial process engineering.

Since World War II, the arts of the chemical analyst have increasingly shifted to measurement of properties that did not require physical segregation of the substance of interest. These "instrumental analysis" methods have largely, but not exclusively, utilized properties not associated with chemical reaction, such as molecular and atomic spectra, nuclear magnetic resonance, and the like.

To the extent such instrumental methods have taken over from wet analysis means within the chemical process industries, a fertile, beneficially close-coupled data flow from the analyst to the process engineer has dried up. Conversely, many of the instrumental methods of analysis in contrast to the often slow and tedious procedures of wet analysis provide a near-instantaneous measurement, which often can be obtained directly from a sample of in-process material. Particularly where a continuous-process technology is involved, the analytical instrument can become part of a sensor-controller linkage for automating suitable portions of the process operation. In some of the more sophisticated process areas, this concept has found important applications. The potentials for more close-coupled process control have also challenged chemical engineers to attempt processing concepts that would not be feasible without such dynamic control capabilities. Finally, most instrumental analysis apparatus produces machine-readable records, or can readily be coupled to recording and computational equipment.

With instrumental process control, there appears to be a possibility that the scientific computer art may ultimately challenge plant managers to assimilate the operational history of a process plant into a sort of "technological memory" capable of defining operating conditions optimal for a given day's mix of raw materials make-up and product-mix demands. The petroleum industry has already demonstrated the economic merit of computer-assisted balancing of process-plant operations to accommodate known feedstock resources and market demands. However, the unusually strong scientific foundations of that process art, rather than cumulative plant operating records, provide most of the necessary data for the petroleum engineer.

<u>Ways Chemical Data are Expressed</u>. Whether chemical data are generated in scientific investigation, through industrial activity, or

by such service professionals as the chemical analyst, the formats for their expression tend to display the substance-property conceptual structure. The substance term tends to reflect molecular identity as much as its purity allows. The property term will reflect the dominant fundamental phenomenon, or a dominant intrinsic attribute that is normally a combination of fundamental phenomena. These terms usually express something measurable, and the data are thus typically numerical. Where such intrinsic environmental parameters as temperature affect the measurement, a numerical expression of the environmental condition or range is part of the data term. With these several determinants requiring identification, graphical and mathematical formats are prized. When they can be made to express a data domain, the need for a record composed of individual measurements has been avoided. Since the reservoir of data-expressable chemical knowledge is enormous, compression of data into such formats is valued particularly by authors, editors, and (when regeneration of the number serving his needs is not tedious or uncertain) users.

It is clear, however, that chemical data are seldom reported or published in a way that irrevocably separates the data from the circumstances surrounding their original measurement. There appears to be a common though unvoiced agreement among authors, editors, and users that the indeterminacies associated with sample purity and property measurement are important qualifying restrictions that must remain associated with published chemical data. At the publication level, the best-regarded handbooks and compilations provide citations to source journal publications. Within the organizations where the data were generated, internal documentation relative to the measurement circumstances is considered an important, usually permanent, record. The existing volume of seldom-used back-up documentation supporting the published chemical data records is undoubtedly enormous.

3. Data Flow

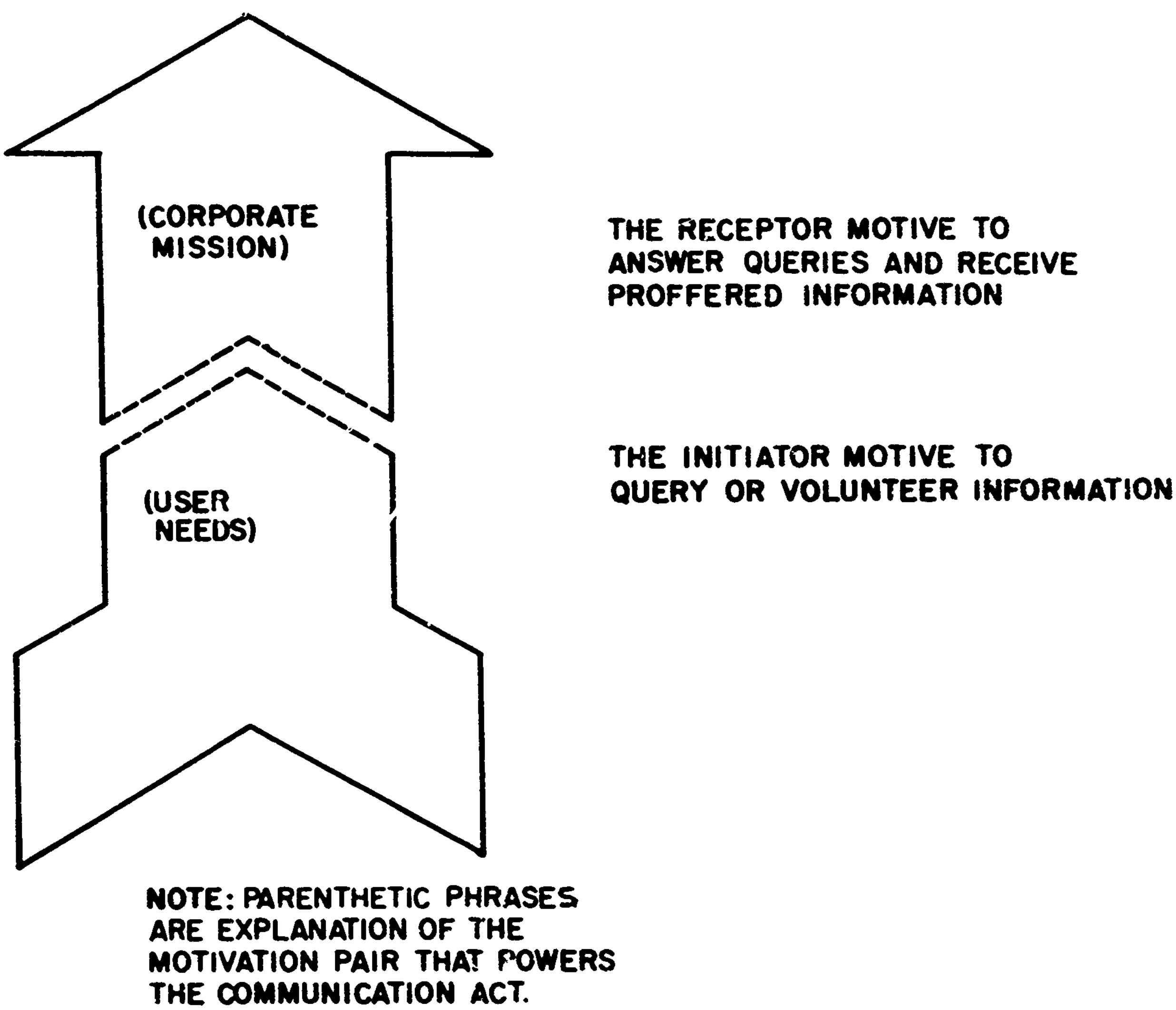
As in any flow process, the flow of chemical data is not fully accounted for without some consideration of the communications forces that drive the specialized world peopled by scientists and technologists. Each of them possesses a personal professional equity, and

most also owe an employee's allegiance and access to technically-motivated institutions. Those scientists and technologists who use chemical data find it enormously abundant, and it is generally accessible if a sufficiently determined effort is made to acquire it. More chemical information is thrust on the typical professional than he can afford to assimilate. Except for data expressive of competitive, proprietary, or other institutionally restricted interests, the world of chemical knowledge might be well considered quite a democratic freemasonry. Under the established bonds of a common professional fraternity, even the neophyte usually can secure sympathetic assistance in meeting his data needs from the most eminent authority, if he has done his homework well.

This lack of an institutional monopoly also has meant that a great variety of motivations are operational in the domain of chemical data generation, aggregation, dissemination, and use. From the systems perspective, the chemical data flow process comprises a largely informal, interconnecting network, much resembling a road map in its provision of many alternate routes to a typical data objective. In fact, the analogy can even be extended to the option of constructing a road directly to the desired destination. This option is genuine, since many types of chemical data can be created or re-created predictably and relatively inexpensively in the user's laboratory.

In this democratic informational environment (which is a general characteristic of the discipline sciences), the existence of chemical data of any type by no means guarantees its actual flow through available communication channels. The communication of information is volitional. Behind each communication act can be found collateral pairs of motivations -- sometimes a chain of them -- that have linked the source of the information to the user. The source-user relationship is essentially contractual; for a consideration, each participant in the communication endeavor has worked to provide something the other desired.

In the discussion that follows of significant chemical data flow relationships between generators, intermediaries, and users, the following symbolic element has been used to represent the motivation sets explaining key communication steps in the flow process:



The Motivation-Pair Symbol

Through this symbolism, the permissive nature of the informal communication system becomes more apparent. The purposive interests of institutions and individuals can be reflected, instead of treating the participants as if they were intellectual automata assigned to designated tasks by a master system designer. From the system conceptualization standpoint, the motivation symbols should be visualized as prime-mover energy inputs that drive the mechanism of the system -- that make it viable.

(It should be noted that even management information systems are incompletely characterized without careful acknowledgement of the prime-mover energy input, even though the motivation-pair relevant to official information systems of a monolithic institution is largely represented by the employment contract. The energy input is organizational, e.g., there is an executive assertion of the desirable course of communication and an executive willingness to pick up the tab for following that course. In such systems, of course, individuals are assigned to designated communication tasks and are expected to be subservient to the system design.)

Probably the most familiar and long-standing communication structure for chemical data is the one that revolves around the scientific society and its journals. The structure depicted in Figure II-D-2 expresses the linkages between the journal author and reader, and the motivational forces that power the journal publications function of the professional society. Several points seem worthy of special note in this flow chart. The most prominent of them is the fact that the individual rather than his institution tends to dominate the flow process.

The journal organization tends to perform a secretariat-support facilitating service that is essentially powered and directed by the publishing needs of authors and the information needs of readers. The most potent power element of communication control within the journalistic mechanics is the reviewer. In the best-regarded journal practice, he is not an editorial employee of the journal, but a professional specialist in the subject of the paper under review. It should be noted also that if the data content of the journal is insufficient to resolve the user's need, the tradition of authorship

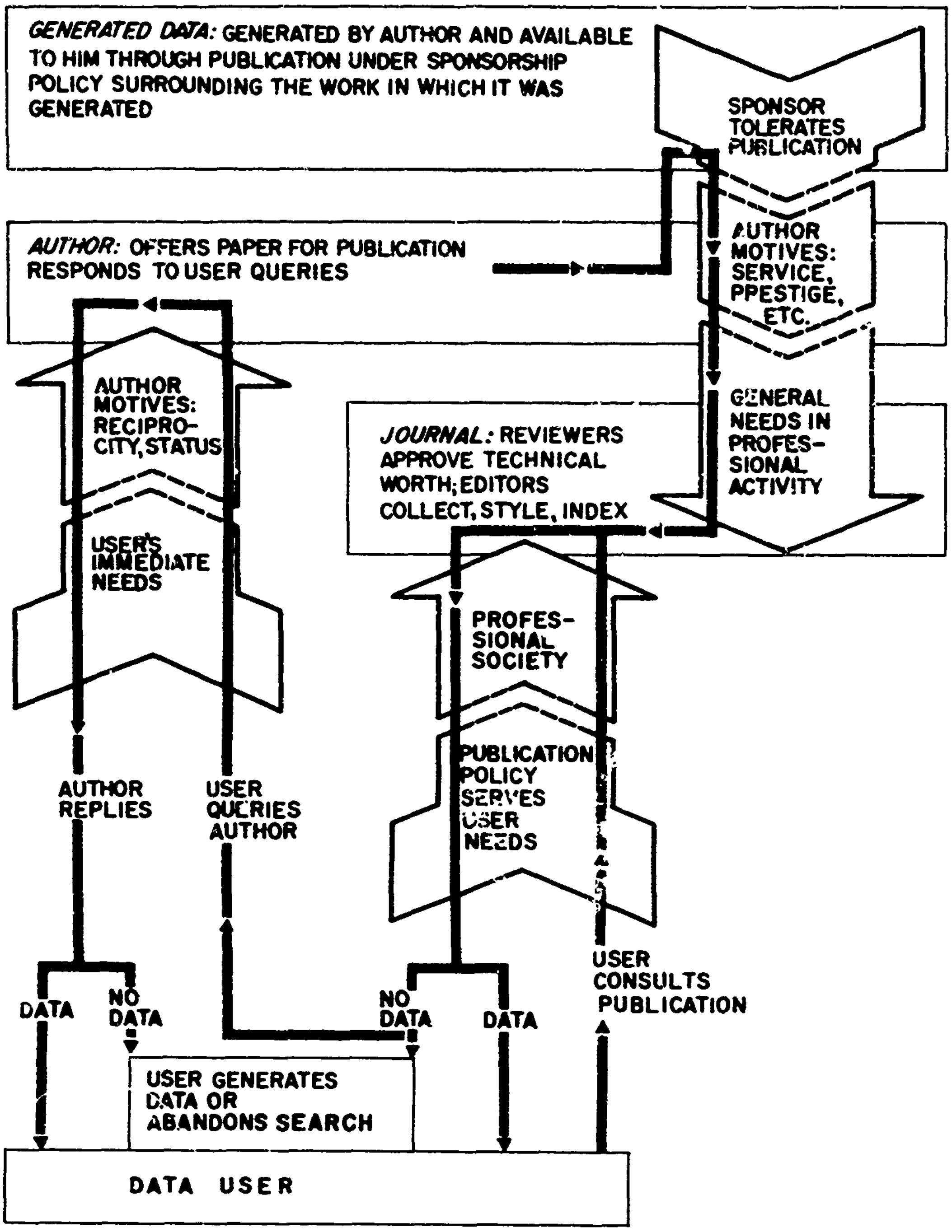


Figure II-D-2 A View of Motivational Forces Acting in the Generation and Flow of Data Through the Mediation of the Scientific Journal

identification facilitates direct communication between author and user. (Secondary publications, e.g., abstract journals, also identify the author. When the user recognizes the need for the author's judgemental aid, or timeliness is an urgent consideration, he is thus given the means to go direct to the author, bypassing the journal completely.)

The direction of the motivation-pair arrow connecting the author and the journal also deserves a brief comment. While it is true that journals appear monthly in the mailbox or on the desk of the subscriber, the significant first communication act of that linkage occurs when the user turns the cover.

The user's options, should he fail to find the data sought, are symbolized by the action box located in juxtaposition to him. The option of abandoning the search can mean that he has decided to seek alternate types of data that may resolve his technical need just as acceptably. The road map to successful technical accomplishment can have as many optional routes as the information channels to a specific piece of data.

The sense of institutional mission expressed by major flow forces in trade-press communication structures is evident in Figure II-D-3. Flows related to the feature-articles and the commercial-literature content are depicted. It will be noted that these flow patterns differ in many respects from that associated with the scientific journal. One of the more striking contrasts exists in the role of the publisher, who displays a multi-coupled relationship with the industrial and trade institutions in his field. Only in the rather lightly invoked communications between the reader and the editor concerning feature articles is there much significant data-seeking communication between professionals functioning as individuals. The remainder of the pattern largely comprises communication between the user and an institution.

A third pattern that is somewhat intermediate of the first two characterizes the commercial publication of handbooks and chemical data compilations (Figure II-D-4). Here, the publisher appears as the specific initiator. However, he displays a major dependancy on

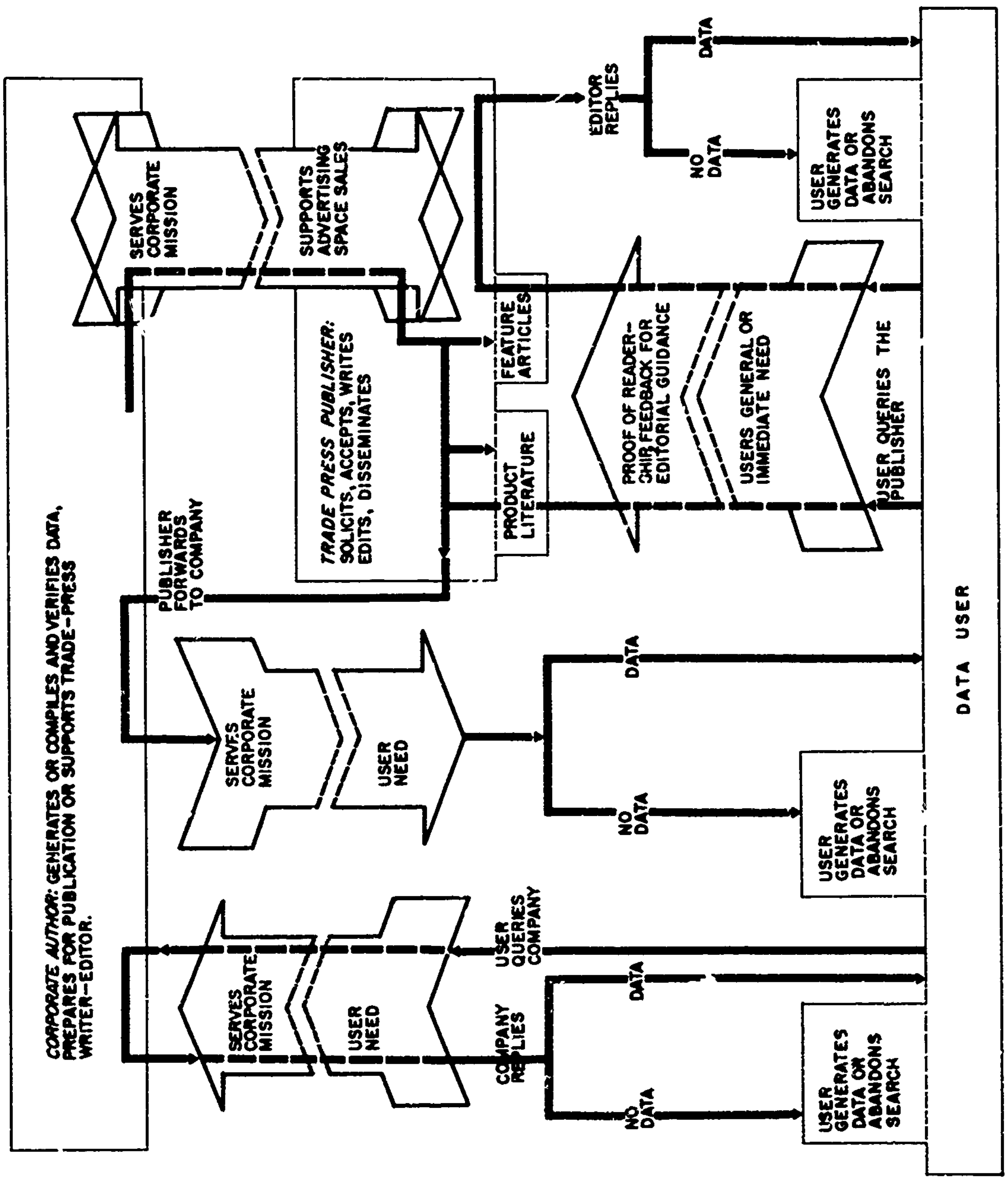


Figure II-D-3 A View of Motivational Forces Acting on the Generation and Flow of Data Through Trade-Press Channels

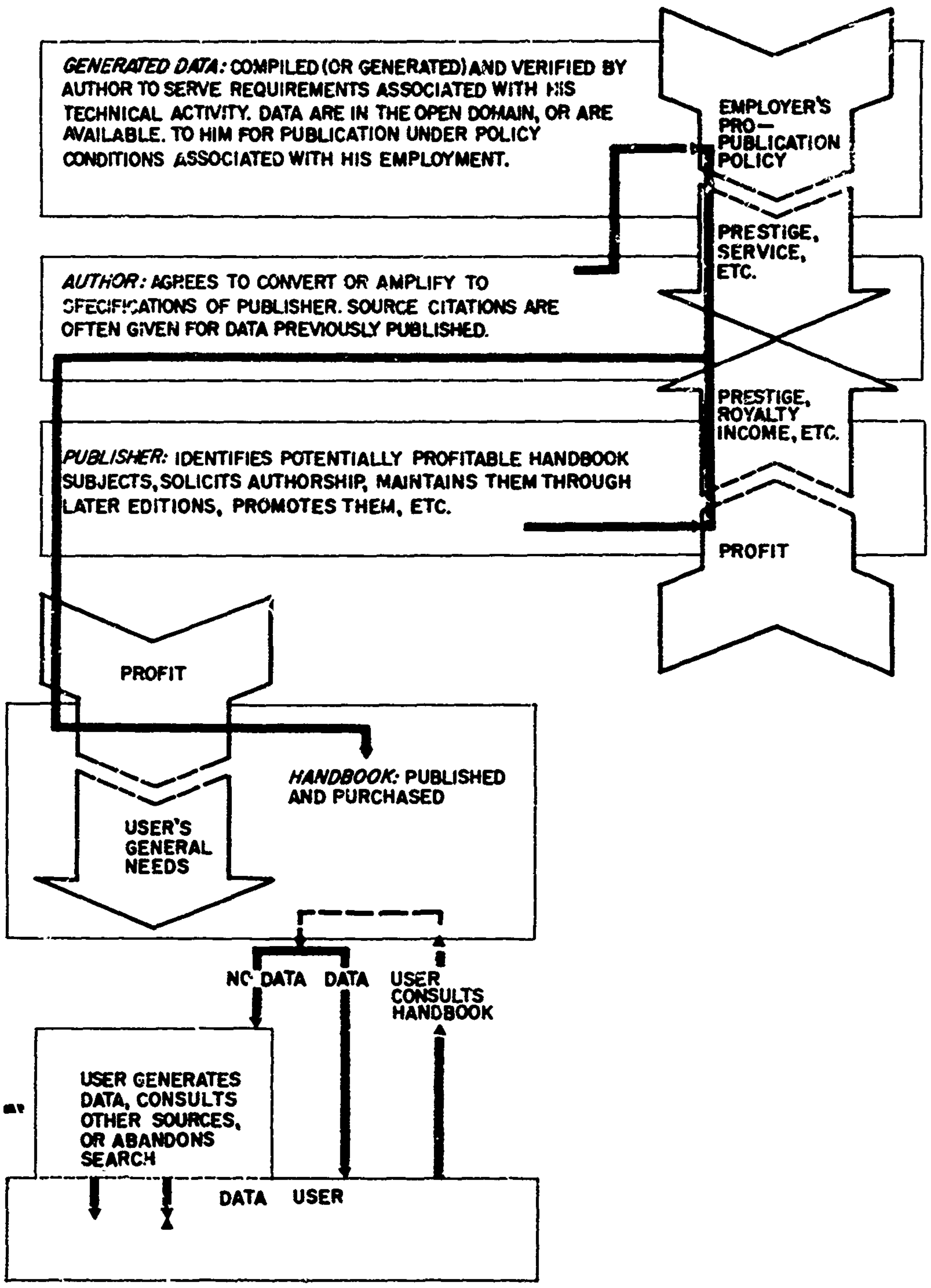


Figure II-D-4 A View of Motivational Factors Associated with the Commercial Publications of Technical Handbooks

finding and establishing viable partnerships with authors possessing access to significant bodies of data, before the actual publication effort takes form. Thereafter, the publisher performs the functions that result in the ultimate delivery of a printed reference document to the bookcase or library of the data user. With rare exceptions, the user-initiated communication activity is restricted to purchase and subsequent consultation of the handbook.

While these three flow patterns are only representative, and not a complete description of the way chemical data flows from generator to user, all three bring out the importance of publication activity as a major ingredient of most data flow processes. A recent study of the chemical data compilation art strongly reinforces this view. (Chemical Data Compilation Analysis Survey, Brunner, R. G., Farris, B. K., Jover, S. I. and Myatt, D. O., Science Communication, Inc., March 20, 1967, AD-652 742.) Most of the activities studied dealt with basic scientific data where one might have thought that interpersonal professional acquaintances might well have formed a communication structure bypassing the scientific literature. Instead, Chemical Abstracts as well as the key journals were prominently represented in the acquisition procedures of most of the compilers. As Figure II-D-5 indicates, except for user-generated data, most chemical data passes through one or more published documents from generator to user via data-compilation organizations. A publication format was even the predominant linkage employed by the most advanced scientific data-compilation groups in communicating with their regular clientele.

The quality of chemical data generated in activities typical of industrial manufacturing, academic research, and process engineering is indicated in Table II-D-3. Perhaps the most significant point to be noted in the tabulation is that there is no sharp segregation of data quality levels from one type of activity to the other. In the university environment, industrial consulting exposes the scientifically-oriented chemist to the pragmatic demands posed by impure materials, and measurement indeterminacies of manufacturing-scale processing. In engineering firms, many of the "measurement" activities comprise calculation of key design, process, and product data for a desired process installation. These calculations typically utilize data from the firm's prior experience with that process, plus raw-

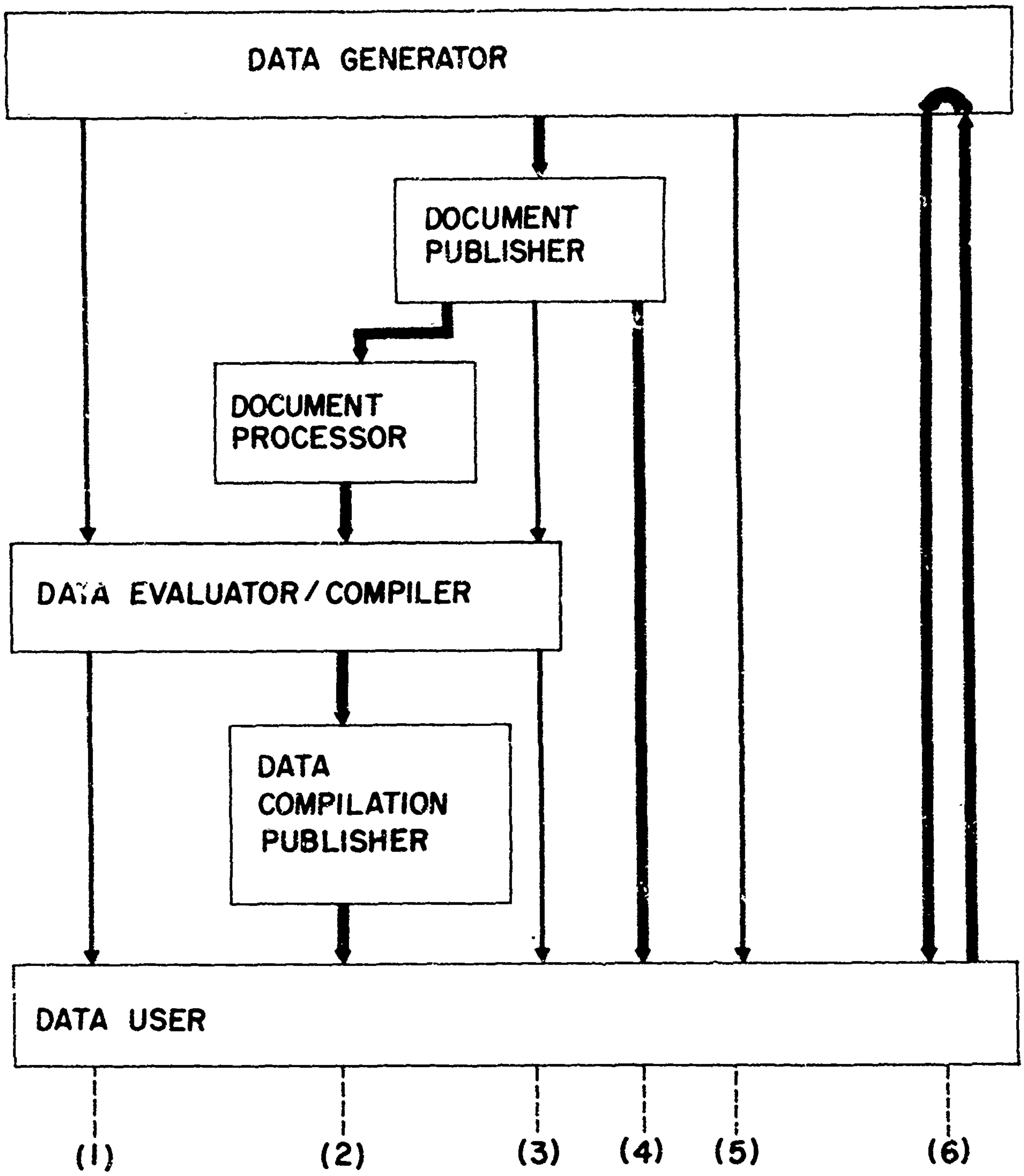


Figure II-D-5 Major Channels of Chemical Data Flow from the Generator to the User

Table II-D-3

Generation and Disposition of Chemical Data
in the Chemical Process Industries,
Academic Activity, and
Process Engineering Firms

	Quality Indicators for Data Most Commonly Generated									
	Substances			Measurement			Disposition			
	A	B	C	A	B	C	A	B	C	D
Industrial Activity										
RDTE Activities:										
Basic or Advanced Research	x	-	-	x	-	-	x	-	x	-
Product R&D	x	x	-	-	x	x	x	-	x	-
Process R&D	x	x	x	x	x	x	-	-	x	x
Product Improvement	-	x	-	-	x	x	x	x	-	x
Process Improvement	-	x	-	-	x	x	-	-	x	-
Application Research	x	x	x	x	x	x	x	x	x	-
Product Characterization	x	x	x	-	x	x	-	x	x	-
Production Activities:										
Process Materials Control	-	x	-	-	x	-	-	-	x	x
Processing Control	-	x	-	-	x	-	-	-	x	x
Product Quality Control	-	x	-	-	x	x	-	x	x	x
Academic Activity										
Undergraduate Research	x	x	-	x	-	-	x	-	x	-
Doctoral and Post-Doctoral Research	x	-	-	x	-	-	x	-	x	-
Consulting Research	x	x	-	x	x	x	-	-	x	-
Process Engineering Activity										
Proposing and Estimating	x	-	-	x	x	-	-	-	x	-
Detailed Designs	-	x	x	-	x	-	-	-	x	-
Testing of Completed Plant:										
Process Materials	-	x	x	-	x	x	-	-	x	x
Process Operations	-	x	x	x	x	x	-	-	x	x

LEGEND

Substances:

- A. Relatively pure substances or carefully analyzed mixtures.
- B. Industrial-specification grade.
- C. Raw ores, variable grade substances, etc.

Measurement:

- A. Research quality; measurement error documented or estimatable.
- B. Industry-standard quality, including specialized methods and equipment common to the industry.
- C. Empirical, ambient, of low accuracy or precision, etc.

Usage:

- A. Original, reduced, or summary data may appear in scientific or trade publications.
- B. Data may appear in company literature or advertisements.
- C. Data preserved in internal files for further technical use.
- D. Data of transient value for process control, product certification, etc.

material and/or product specifications from the client, basic chemical data from the journal literature, and sometimes laboratory data from a few key experiments. This intermingling of data types and practices indicates that chemists and chemical engineers can shift rather easily from scientific to pragmatic levels of manipulation as their technical activity can take advantage of different levels of technical sophistication.

The data generated and disseminated in connection with the commercial activity of the chemical process industries is voluminous, extremely important to the technologically-oriented user, and generally of technical quality well suited to its principal usages. Figure II-D-6 illustrates the sectors of a chemical company's activity cycle where data of various types are generated and disseminated. In Figure II-D-7, the specimens of product-promotion literature displayed show the high technical quality and customized level of data-communication linkage that chemical companies are prepared to establish at the first point of contact with the user. Partly because they are so readily assembled, commercial chemical data collections fitting the individual professional's specialization form an important fraction of his personal files. The same, of course, is true for the data file resources of technical organizations. Within the past few years, vendor-data service activities have assumed an increasingly important role in commercial-data communication systems. This role is analogous to that performed by the abstracting-indexing secondary publications in the scientific literature. Vendor-data systems are comparable in some respects to the long-established commercial-product catalogs. However, they use economical photocopy-microform update approaches that have avoided the built-in obsolescence of the bound book, a limitation that is particularly detrimental for information that fluctuates as new products or grades are introduced and old products discontinued. There appears to be a current trend for large publishers to acquire independent vendor data service firms, or utilize byproducts of their trade-journal activity to establish such services as part of their communications-service structure.

4. <u>Principal Problems and Prospects for Chemical Data Systems.</u> The major problems confronting the designer of nationally significant chemical data systems generally associate with the large dimensions he must contend with. It is hardly an exaggeration to say that chemical

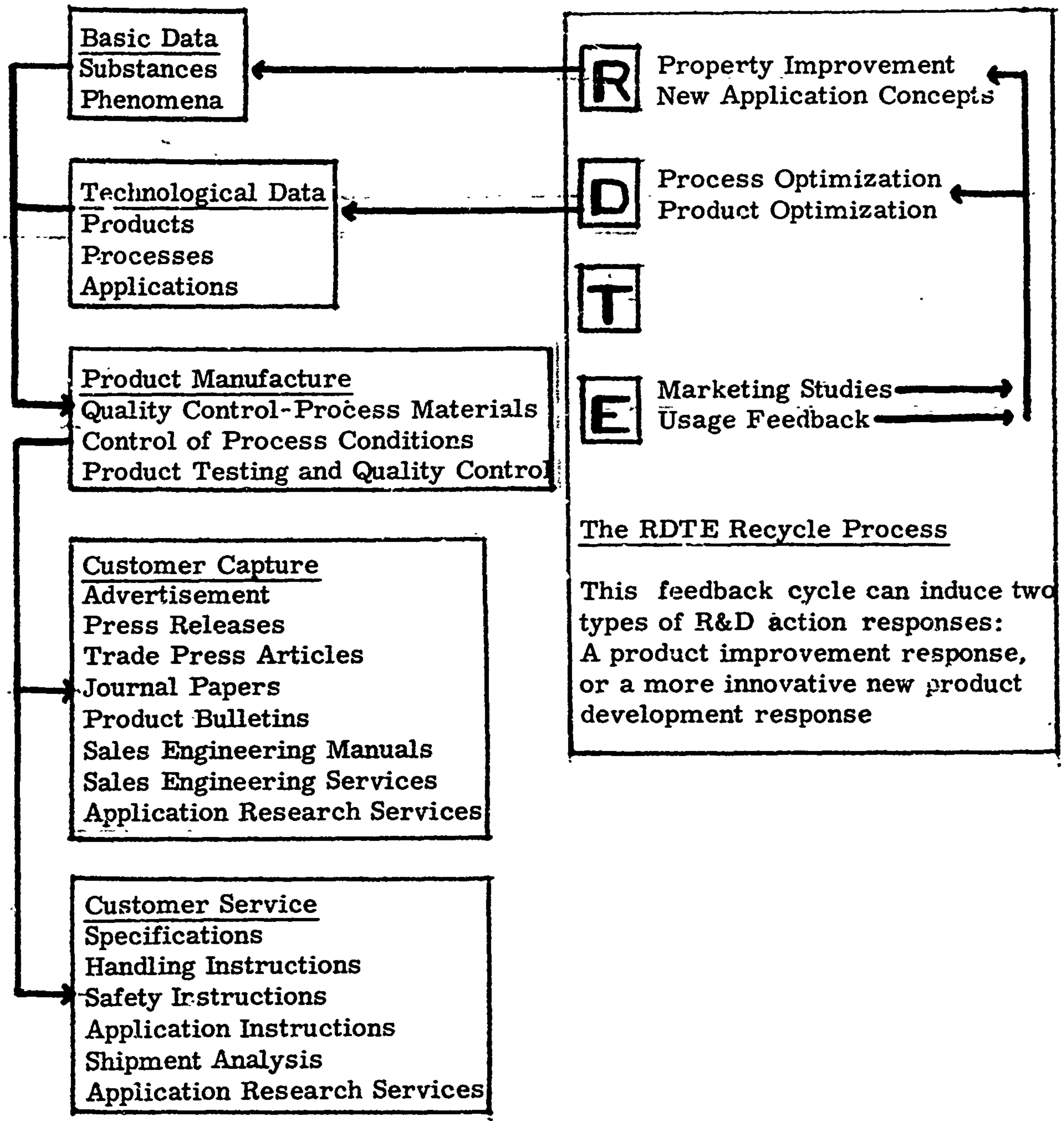


Figure II-D-6. Some Functions Where Technical Data Are Generated, Communicated, or Used in the Activities of a Chemical Manufacturer

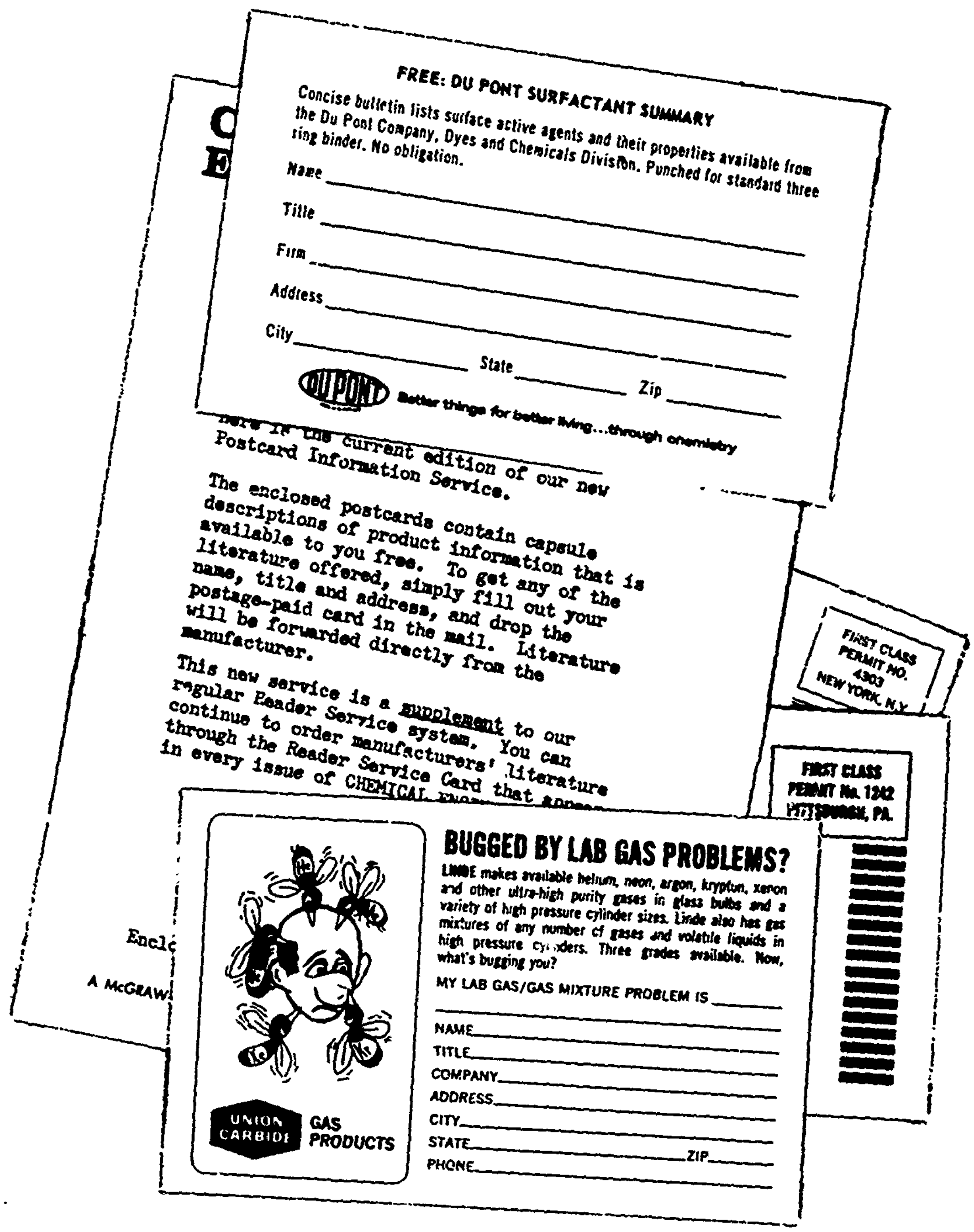


Figure II-D-7 <u>Examples of Methods Used to Communicate Commercial Chemical Data</u>

data of one class or another are used by virtually the total population of modern scientists and technologists. The rate of generation of chemical data is a continually growing flood. New and valuable ways to utilize chemical data proliferate and advance in sophistication.

The most exciting opportunity held forth to the designer is the powerfully articulate, systematic language of substance identity and property. Chemical data can be readily codified, and individuals in modern societies are taught its essentials as an important ingredient of their cultural heritage. In storage-retrieval systems, the interface between system and user poses few problems of any intellectual consequence.

The two major attributes of the chemical data resource -- its great size and its exceptionally strong structure -- provide perspective from which to examine the data systems and institutions important to chemical activity today, and what might logically be foreseen or sought for the future.

In today's chemical activity environment, the scientific societies and other institutions principally dedicated to the conservation and dissemination of chemical science show signs of considerable stress. Parts of this stress can be traced to their service charters, which traditionally express the obligation to accommodate all scientifically meritorious material within their charter scope. These charters have begun to be losing propositions economically, with the methods of information conservation, dissemination, and financing that have been traditional to the scientific society. The result has been a recent era of apprehensive and uncertain experimentation with new revenue sources such as the page charge for scientific papers, contract or grant income from the Government for secondary-journal coverage of foreign-language literature, and similar outside underwriting of the development costs of newer dissemination tools and methods. In the publication function, budgetary pressure has tended to erode the "comprehensive coverage" concept in practice, undermining editorial vigor in fulfilling the declared charter, or shifting editorial initiative toward specialized service concepts that would not produce financial disaster if pursued with full diligence and technical success.

A second major source of the stress on major chemistry-conserving institutions originates with some of the major chemistry-utilizing institutions. The forces behind this source have found their voice only recently -- essentially, since Sputnik. The largest of these institutions, the Federal government, is a hermaphrodite in its functional relationship to technical information systems. Its broad charter obligates it to create chemical information systems not otherwise existing, should they be called for in the general public interest. As a user -- for chemistry is important to most of the Government's large and diverse technology programs -- its own operational management requires chemical information systems adequate for the effective performance of their technical missions. The spur of this latter obligation seems to account primarily for the raising of the Government's voice. The role of the scientific society will clearly be influenced if major operational Government systems, available to the public, and disseminating discipline-science information, are established.

Among the world's scientific societies, the American Chemical Society has been one of the most successful in maintaining its ancient charter and traditional operational patterns as the knowledge special to its field has grown. It is a strong organization, with no evidence that the stresses of the present era threaten its collapse, or are likely to precipitate ill-considered decisions resulting from a panic psychology among its managers. The Society is looked on by Government managers as a responsible and knowledgeable interpreter of chemistry in a modern world. These circumstances suggest that a review of the significant recent interactions between the ACS and the Government provide an instructive model of the contemporary problems and issues associated with chemical data systems . . . and to some considerable degree, of discipline-oriented data systems generally.

The bulk of the ACS - U. S. interactions have revolved around the Chemical Abstracts Service. Over the 60-year period since CA's establishment, its original comprehensive coverage charter has been sustained. To an impressive degree, CA continues today to be "the key to the World's chemical literature". Chemical Abstracts is a resource known to and usable by all chemically literate scientists

and technologists. The value of the resource would drop abruptly if its "comprehensive coverage" charter were abandoned. No other existing resource provides an alternate. The Society's financial policy for CA has required its operating costs to be met by subscription and service revenues. Through periodic price increases that have become substantial, it has maintained a break-even cost history to the present time.

There have been significant government needs for chemical information processing that CA is well-equippped technically to perform. Some exploratory and experimental uses of the CA resource have been made by the Department of Defense and the National Institutes of Health, which are accustomed to using contractors for such services. The ACS response has rather clearly displayed a desire to maintain an operation that is not dependent on government contract or grant income. However, the Society is cooperative in providing contract services that do not jeopardize its independence or alter its basic institutional charter. The current working relationships largely seem to comprise contractual services in which CA "works up" material to be incorporated in specialized Federal operational systems such as the National Library of Medicine. There are few if any instances where CA is coupled operationally to a government system to any greater degree than it will arrange with any user of the CA service. Furthermore, CA's special contract services appear to be usually associated with chemical information products tending to conform the agency's system to the Society's, rather than the reverse.

A second major U. S. - ACS interaction has also developed which displays recognition of the continuing national need for strong chemical information systems. At present, this interaction is expressed principally through the instrumentality of the Chemical Information Program administered by the National Science Foundation. CIP is establishing, through technical and system-analysis grants and contracts, the design directives for a national chemical information system. The intent is to develop a system that can provide technically advantageous coupling -- at the management, special-service, or regular operational levels as appropriate -- between significant generators, disseminators, and users of chemical information. The Chemical Information Program is the most advanced operationally of several similar programs through which the Foundation

is endeavoring to develop a structure of national-scale, discipline-based technical information systems.

CIP has approached the ACS with the proposal that the Society become the "chosen instrument" to collaborate in the development and ultimately to become the operational focal point of the national chemical information system. Federal funding support is presumed for the developmental and, if necessary, the operational phase. In September 1967, the Society responded by designating a ranking staff member to help develop a fully-considered position concerning this important proposal. This appointment is generally interpreted to mean that the Society intends to accept additional roles judged important in the national interest, within the limits of prudence.

Since the late '50's, the ACS has also received Federal support for research and operational-level applications of modern data-processing technology to the handling of chemical and chemical engineering information. Virtually all of this work has been directed toward upgrading or superseding the Society's existing products and information services. The great bulk of the effort has focused on conversion of the Chemical Abstracts operation to a computer-based, photocopy-printout technology, and to develop a computer-based substance-identity registry system. The registry system is designed for machine-searching for submolecular structures of interest, as well as accommodating all scientific and empirical synonyms for the substance. Approximately 800,000 substance identities have been entered into the Registry system to date and 4,000 additional entries are being made weekly.

Historically, this R&D work preceded and undoubtedly was an important influence in the gestation of the Chemical Information Program, which was formalized about three years ago. In 1967, a CIP-supported contractor developed a plan for a national chemical information system, to be fully developed by 1972, at a total cost of approximately 18 million dollars.

The principal CIP activity continues to be the technical development of the CA computer methodology. In addition, CIP continues to sponsor some exploratory and pioneering development studies by non-ACS contractors. This situation appears likely to continue until NSF

receives some definitive answer from ACS on the "chosen instrument" overture. In the meantime, CA's NSF-supported processing developments advance its technological competence, but do not affect its continuing operational independence.

The Chemical Information Program is broad, and explicitly anticipates provision of data systems in the total system plan. As mentioned earlier, a study of the chemical data compilation art had demonstrated the central role of the scientific journals in general and Chemical Abstracts in particular, as screening sources utilized for acquisition of newly published data. One key recommendation arising from that study was that a feasibility demonstration be conducted of property-indexing strategies effectively compatible with CA's present highly developed, computer-searchable substance-identity characterization. If implemented to the operational level in abstracting- indexing publication practice, a major improvement in the accessibility of the data content of the chemical publication system would result. A second key product of the chemical data study was a suggested schematic (Figure II-D-8) relating the chemical data subsystem to the publication subsystem of a national chemical information system. The flows on this schematic reflect the ultimate desirability of developing direct-linkage traditions between data-generators and data-oriented subsystem elements. An essentially decentralized data sub-system was recommended. This was considered particularly desirable in order to conserve information critically significant to highly specialized subgroups of data generators and users.

From this necessarily sketchy review of the U. S. - ACS chemical information story, it is evident that a rather substantial framework of thought , action, and accomplishment already exists on the national scene that has significant application for any future chemical and chemical engineering data systems. Beside the ACS story, the story of the Army's Chemical Information and Data System, the Standard Reference Data System of the National Bureau of Standards, and the subject-oriented national technical information centers under development or in operation under the sponsorship of various Federal agencies also contain many organized data collections relevant to chemistry and chemical engineering. The chemical engineering literature burgeons with computer-based techniques for manipulating data

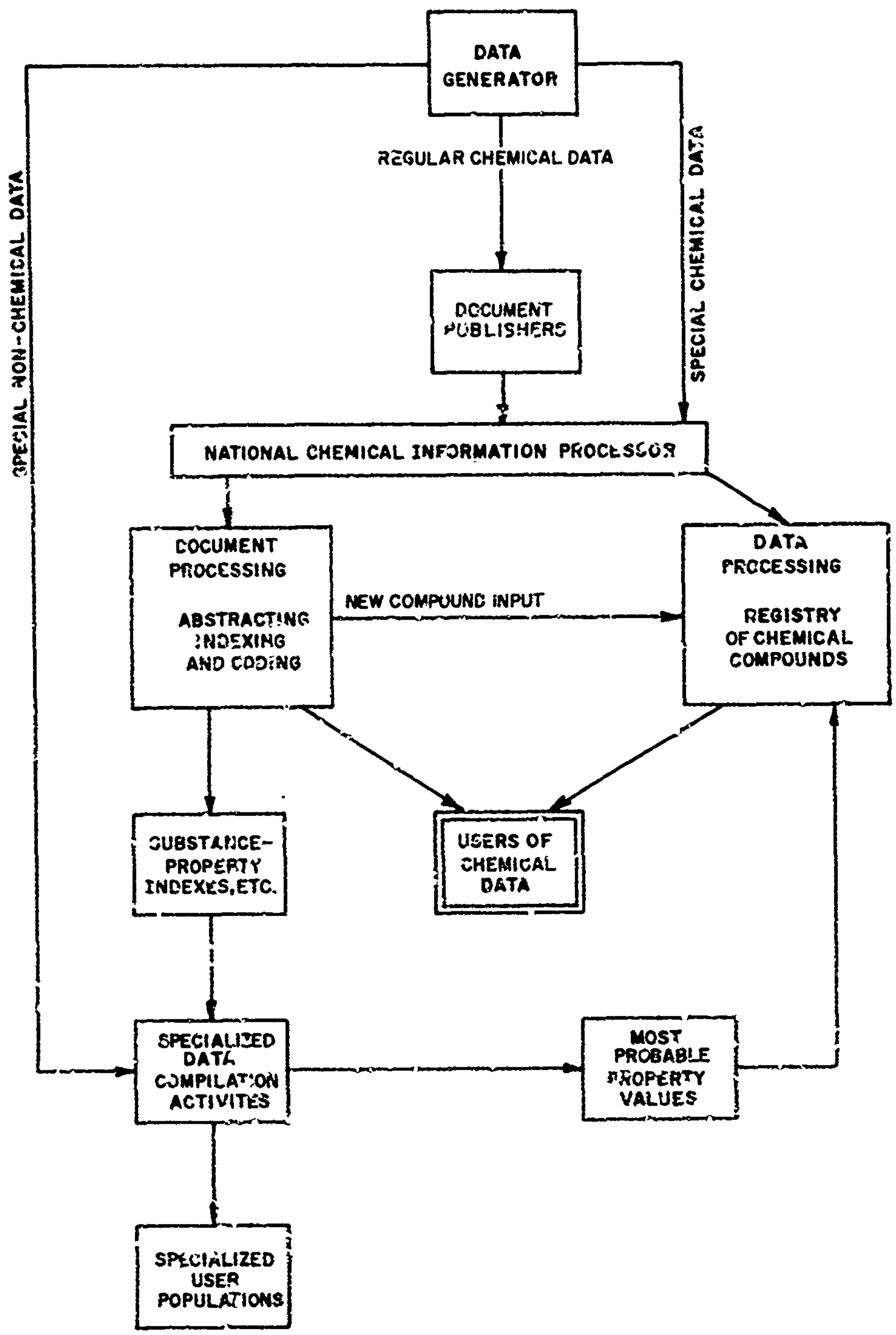


Figure II-D-8 Schematic of a Proposed National Chemical Information System Incorporating Specialized Data Compilation Activities

in increasingly powerful ways for applications ranging from process design, to computer methods for estimating physical properties that may be required to use computers for process design.

Beyond the professional society and government institutions, the industrial and commercial organizations have dealt with their chemical data problems and resources pretty much as one more factor associated with the cost of doing business. Only rarely, as in the earlier-mentioned case of the American Petroleum Institute's sustained sponsorship of basic chemical data activity in hydrocarbons, are industry-wide data efforts undertaken to maintain or advance the quality of the general technology. Where external accountabilities are involved, such as product-testing standards used by customers, or government approval for product sales, jointly supported data efforts occur in trade associations and such institutions as the American Society for Testing and Materials.

There are major accumulations of chemical data in the industrial and commercial sector, much of it of high quality. There are competent management structures associated with these resources. If these managements can be effectively motivated to participate in national chemical data systems, an effective operational input from the commercial sector might be achieved essentially at the click of a policy switch. For many of these institutions, no more inducement may prove necessary than the increment of processing and administrative cost needed to couple the operational linkages.

In contrast to this condition of institutional interaction, the individual chemist and chemical engineer currently seems to be just that -- very much an individual -- in his relationship to his information and data environment. It is as if he recognizes that the job of the structuring of the data resource for orderly or formal accessability is beyond the capacity of the professional scientist, whether he acts singly, or in concert through the scientific societies he has created as vessels for joint efforts with his peers. The professional data resources he uses in serving his vocational obligations are likewise largely a matter of personal decision. Today, there is virtually no recognizable "recommended practice" tradition for the quality of data resource utilized by the employed chemist and chemical engineer.

Until some "superhighway" structure is created that establishes institutionalized access paths to the rich chemical data resources now accessible only through informal means, the data practices of chemists and chemical engineers will probably continue to be a matter of self-discipline and personal style. Management of technology innovation programs will continue to be (or perhaps increasingly become) a matter of leadership and support of talent rather than a controllable, planned exploitation of human and scientific capability to achieve defined goals.

SELECTED BIBLIOGRAPHY

(1) An Overview on Worldwide Chemical Information Facilities and Resources, prepared for ICSU-UNESCO Joint Study on the Communication of Scientific Information and on the Feasibility of a Worldwide Science Information System. Office of Science Information Services, NSF (1967).

(2) Reviews of Data on Science Resources, 1966, NSF 66-34, National Science Foundation (1966).

(3) Chemical Statistics Handbook, 6th Edition, Manufacturing Chemists Association (1966).

(4) Chemical Data Compilation Analysis Survey, Brunner, R. G., Farris, B. K., Jover, S. I. and Myatt, D. O., Science Communication, Inc., March 20, 1967, A. D. 652 742, Federal Clearinghouse for Scientific and Technical Information.

(5) AIChE Program for the Computer Estimation of Physical Properties, American Institute of Chemical Engineers (1965).

E. Agriculture and Food Technology

1. Introduction

For the purpose of this study, the field of agriculture and food technology is defined as the science and technology associated with protection and production of farm and forest resources, and the associated technology of food product production. It includes both plant and animal resources, but devotes minimal attention to the toxicological and medical aspects which are treated in the Pharmacology and Biomedical sections of this report.

Scientific and technical activity in the field of agriculture is dominated by three organizational groups: the U.S. Department of Agriculture (USDA), the State Agricultural Experiment Stations (SAES), and private farming industry. To the extent that USDA and SAES are involved in supporting the production of raw food materials, they are also closely involved in the scientific and technical activities in the field of food technology. Apart from these involvements, the principal organizational entities involved are the food products industry and the Food and Drug Administration.

This section provides a brief characterization of the broad classes of data of importance in the fields of agriculture and food technology, and in broad terms, traces the flow of data from production of raw materials to the manufacture of food products. The enormity of the associated scientific and technical effort is indicated by the dollar volume and manpower investment therewith associated. Funding for agricultural research and development is estimated by the USDA to exceed eight hundred million dollars per year; with industry providing 55 percent, USDA providing 25 percent, and state agencies providing some 14 percent. The goals and corresponding levels of funding (for FY 1965) are shown in Table II-E-1, and the scientific manpower requirements projected for meeting these goals for FY 1972 are shown in Table II-E-2.

TABLE II-E-1. AGRICULTURAL RESEARCH AND DEVELOPMENT PROGRAM BY GOAL IN FY 1965

Source: "A National Program of Research for Agriculture", USDA, 1966.

Goal	USDA		SAES		Total USDA-SAES		Industry		Total USDA, SAES, & Industry	
	Million $	%	Million $	%	Million $	%	Million $	%	Million $	%
(1) Resource conservation and use	28.6	17	19.9	9	48.5	12	17.0	4	65.5	8
(2) Protection of forests, crops, and livestock	43.8	26	46.5	21	90.3	23	115.0	25	205.3	24
(3) Efficient production of farm and forest products	25.2	15	98.4	43	123.6	31	115.0	25	238.6	28
(4) Product development and quality	39.0	23	26.6	12	65.6	17	160.0	35	225.6	26
(5) Efficiency in the marketing system	9.4	6	7.2	3	16.6	4	15.0	3	31.6	4
(6) Expand export markets and assist developing countries	1.1c	1	0.5c	a	1.6c	a	8.0	2	9.6	1
(7) Consumer health, nutrition, and well-being	11.1	7	9.5	4	20.6	5	22.5	5	43.1	5
(8) Raise level of living of rural people	3.1	2	3.4	2	6.5	2	1.5	a	8.0	1
(9) Improve community services and environment	5.6	3	14.7	6	20.3	5	6.0	1	26.3	3
Total	166.9	100	226.7	100	393.2	100	460.0	100	853.6	100

a - Less than 0.5 percent.
b - No observations in the sample.
c - Does not include funds to support USDA-SAES research workers located in foreign countries working on problems of foreign country.
d - Based on only one observation in the sample.

TABLE II-E-2. SCIENTIFIC MANPOWER REQUIREMENTS TO ATTAIN AGRICULTURAL RESEARCH GOALS

Research Goal	Seven Year Growth (%)	% of Program in 1972	Scientist Manpower (man-years)	
			1972 (est.)	1965
(1) Resource conservation and use	31	12	1,750	1,300
(2) Protection of forests, crops, and livestock	37	21	3,000	2,200
(3) Efficient production of farm and forest products	22	26	3,750	3,200
(4) Product development and quality	18	15	2,250	1,750
(5) Efficiency in the marketing system	21	5	900	750
(6) Expand export markets and assist developing countries	267	5	700	250
(7) Consumer health, nutrition, and well-being	60	5	800	500
(8) Raise level of living of rural people	135	4	575	250
(9) Improve community services and environment	86	7	1,250	750

Source: "A National Program of Research for Agriculture", USDA, 1966.

The distribution of scientific and technical effort, and therefore, the associated distribution of data activities in the attainment of the nine goals of the national agricultural program, are in accordance with the distribution of disciplines involved. Table II-E-3 summarizes this distribution of effort for FY 1965. The subsection to follow characterizes the data activities by class of data.

TABLE II-E-3. AGRICULTURAL RESEARCH AND DEVELOPMENT BY FIELD OF SCIENCE IN FISCAL YEAR 1965

Field of science	USDA		SAES		Industry		Total	
	$ Million	%	$ Million	%	$ Million	%	$ Million	%
Biological	91.6	55	176.3	78	142.6	31	410.5	48
Physical	59.1	35	32.2	14	308.2	67	399.5	47
Social	16.1	10	18.1	8	9.2	2	43.4	5
Total program	166.8		226.6		460.0		853.4	

Source: "A National Program of Research for Agriculture", USDA, Oct. 1966.

2. Data Characteristics

The data associated with the field of agriculture and food technology may be grouped into categories closely associated with the pattern of transition from raw food production to food product manufacture and marketing. Table II-E-4 summarizes the classes of data which follow this pattern.

TABLE II-E-4. TYPICAL CLASSES OF AGRICULTURAL DATA

Scientific or Technical Activity	Related Data Classes
(1) Natural Resource Conservation and Utilization	Soil Resource Data - soil characteristics (physical, chemical, biological) - soil response to environment and utilization data
	Land Management Data - systems, techniques, and equipment performance data - mathematical models for land performance prediction
	Environmental Data - weather and climate forecasts - biological and physical consequences of environment
	Ecological Data - forest and plant physiological and ecology data - forest range data - rodent, insect, and plant pathology data
(2) Forest Crop and Livestock Protection and Production	Insect, Disease, Vermin, and Weed Effect Data
	Fire and Environmental Hazard Data
	Pesticide Performance Data

TABLE II-E-4. TYPICAL CLASSES OF AGRICULTURAL DATA (Cont'd)

Scientific or Technical Activity	Related Data Classes
(3) Farm and Forest Production	Tree Reproduction and Growth Data
	Forestry Systems and Techniques Performance Data
	Fruit and Vegetable Crop Production Data - genetic and plant physiology data - breeding performance data - production equipment performance data - crop management data
	Livestock and Poultry Data - feed efficiency data - genetic and environmental response data - management technique performance data
	Equipment Engineering Data
(4) Product and Product-Quality Development	Forest Products Data - tree anatomy data - effects of environment on chemical, physical, and structural properties of wood - pulp yield data - pulping-process performance data
	Fruit and Vegetable Data - market research data - genetic, chemical and environmental effects on product quality - production quality data

TABLE II-E-4. TYPICAL CLASSES OF AGRICULTURAL DATA (Cont'd)

Scientific or Technical Activity	Related Data Classes
(4) Product and Product-Quality Development (Cont'd)	Textile Products Data Animal Products Data - market research data - meat, milk, and egg product quality data - wool, hide, skin, and animal fat data
(5) Marketing & Management (timber, fruits, and vegetables, field crops, livestock)	Product Quality Data Production-Improvement Techniques Data
(6) Export and Foreign Technical Assistance	Product Utilization Data Foreign Farming Conditions Data
(7) Health and Nutrition	Human Nutrition Requests Data Pesticide Residue Data Microorganism Occurrence and Effect Data Home Economics Data
(8) Food Processing	Raw Material Processing Data Conversion Processing Data Packaging Operations Data Process Development Data

Within each class of data shown in this table, there are broad and often complex subclasses which associate several scientific technical functions and activities. An example of an array of subclasses within a specific data class (those associated with food additives) is shown in Table II-E-5. Numerous interrelated bio-medical, chemical, and pure technical art functions are associated with the utilization of these and the many other subclasses of data in this large field. Therefore, characterization of the data is approached from a largely non-discipline oriented standpoint. From a functional standpoint, the data may be classed according to six primary groups:

(1) Research data associated with natural resources management;

(2) Operational data associated with natural resources management;

(3) Research data associated with farming and forestry management;

(4) Operations data associated with farming and forestry management;

(5) Research and development data associated with food handling and processing; and

(6) Operational data associated with food handling and processing.

As in other fields of science and technology, wherein highly ordered disciplines such as chemistry and physics are extensively involved, the research activities are based on generation of data in support of, or for use of, well established rationalizations and theories. Moreover, the operational activities are based on a well-established empirical data base embodied in the experience of practicing personnel. Most of the research data are contained in journal articles, technical society papers, technical reports, and proprietary archives. There are few formal data efforts in the agriculture field, and most of the data are either descriptive in form or are cast in a descriptive context. Access to the data is for the most part through the literature, except in instances where Federal control or other mission-oriented functions are performed, in which cases the data are sometimes extracted and

TABLE II-E-5. CLASSES OF FOOD ADDITIVES AND ASSOCIATED FOOD QUALITY DATA

Food Additive	Data
Acidulants	acidity
Aerating agents	gas content
Desiccants	caking tendency
Anti-oxidants	deterioration or rancidity
Bleaching agents	color
Buffering agents	pH
Clarifying agents	turbidity
Clouding agents	turbidity
Coating agents	glaze properties
Coloring agents	color
Conditioners	dough elasticity
Surfactants	dispersion stability
Enzymes	hydrolysis number
Flavoring agents	flavor, aroma
Foam regulators	foaming tendency
Hydrolytic agents	molecular cleavage tendency
Leavening agents	fermentation power
Maturing agents	aging quality
Nutrients	mineral and vitamin content
Preservatives	spoilage rate
Sequestrants	trapping capability
Sweetening agents	sweetness
Texturizing agents	firmness
Thickening agents	viscosity or thickness

stored in alphanumerical form. Because of this mode of storage, retrieval, and dissemination, data are available to the user through traditional library and publication access channels, and the data user must be generally familiar with the subject area if he is to rapidly obtain data.

Partly because of the restricted rate of data flow imposed by this traditional mode of operation, the obsolescence rate of data in the agriculture and food technology field is quite slow, except in the commercial sectors where competitive forces drive data flow. New research data may take as long as five years to reach the practicing forestry manager or farmer; while new food additive data or packaging data are likely to influence food processing operations as soon as pertinent Federal regulations permit.

3. Data Flow

In the field of agriculture and food technology, data flows according to the traditional modes established for most science/technology activities wherein there is a mixture of discipline-research, mission-developmental and applications activities. Data flow may therefore be characterized according to these three modes of science/technology activity. The three associated classes of data sources are illustrated in Table II-E-6 for a specific area of activity.

- Discipline/Research Data Activity is based primarily on use of the literature. Federally and state sponsored research activities use the literature to report and retrieve research findings. Project Able (see bibliography) lists some 150 journals, abstract publications and periodical bibliographies which are widely used.

- Mission/Developmental Data Activity, including equipment, chemical agent and farming technique development, as well as land or forestry project management use the research literature as well as product vendor data, informal communication and prototype performance data

TABLE II-E-5 CLASSES OF DATA ASSOCIATED WITH FOOD PROCESSING OPERATIONS

Food Processing Operation	Typical Data	Typical Data Sources
Raw Material Preparation	performance data for cleaning, separating, draining, trimming, peeling, dehusking, silking, cutting, shelling, stemming, pitting, filtering, extracting, centrifugal equipment and processes.	equipment manufacturers, trade publication articles.
	raw material specifications.	raw material suppliers, proprietary sources, and the FDA.
	raw material quality control.	internal operations.
Raw Material Conversion	performance data for size reduction, mixing, blending, sterilization, cooling, freezing, antibiotic treatment, crystallization, coating, fermentation, pickling, curing, ageing, smoking, deodorizing, hydrogenating, puffing, whipping, deaeration, emulsifying, and homogenizing equipment and processes.	equipment manufacturers, trade publication articles.
	product standards.	corporate management, and the FDA.
	process control standards and data, and product quality control data.	internal laboratories.
Food Packaging	performance data for feeding, filling, closing, labelling, wrapping, and coding equipment and processes.	packaging equipment manufacturers, and packing operations management.
	packaging materials standards and quality control data.	corporate management.
	packaging standards.	FDA, corporate management.
	packaging operations standards and package quality control data.	corporate management, and internal laboratories.
Food Process Development and Modification	new raw material data.	raw materials suppliers, USDA publications, internal testing laboratories.
	new conversion-process data.	conversion equipment suppliers, internal laboratories, IFT publications, trade journals.
	new packaging - data	packaging materials and equipment suppliers, trade journals, internal laboratories.
	new standards and regulations (such as shelf life, bacterial count).	FDA.

- Applications Data Activity utilizes vendor data, agricultural test laboratory results, and other highly refined data available usually from equipment or material suppliers, U. S. D. A. field representatives, and Federal specifications.

4. Representative Problems

Of all the many data management problems that seem to prevail in this field, two stand out as most prominent: (1) Flow of research results to developmental and applications activities takes from three to five years because of the traditional publication mode of data flow; and (2) Increasing regulations by F. D. A. are motivating development of data management activities that protect proprietary interest rather than promoting advancement of the agricultural arts and sciences. A major study of the data activities in the agriculture field would identify the more specific problems, and indicate the specific data management and data system requirements to overcome these specific problems in the field.

SELECTED BIBLIOGRAPHY

1. List of Available Publications--U.S.D.A. List No. 11.

2. Proceedings of the Annual Meeting of the American Society of Agricultural Engineers, June 1967.

3. Proceedings of the Annual Meeting of the Institute of Food Technology, May 1967.

4. List of Agronomy, Crop and Soil Publications, American Society of Agronomy and the Crop Science Society.

5. A National Program of Research for Agriculture, U.S.D.A., October 1966.

6. Agriculture Biological Literature Exploitation, Report of Task Force, U.S.D.A., March 1965.

Science Communication
Washington, D. C. 200 07

COSATI Data Activities Study
Final Report - F44620-67-C-0022

30 April 1968

F. Biomedical Sciences

1. Introduction

For this study, biomedical science is defined as the art and science of preventing and treating hazards of human life and health. It includes the data activity associated with both basic medical sciences and clinical activities, but only covers those relating to veterinary medicine to a limited extent. Section II-E on Agriculture and Food Technology in this volume covered the data activities associated with human nutrition and animal husbandry. Section II-G of this volume covers the field of pharmacology.

The purpose of this section dealing with data activities in the biomedical sciences is to describe the characteristics, flow and problems associated with biomedical data. At the outset, it is important to note that the superior medical care available in the United States is due in some measure to the data and information transferred from biomedical research to the practicing physician. But the increase of technical information in clinical medicine and the biomedical sciences, particularly in the past 20 years, has been most prodigious. It follows that there have developed concomitant problems in the storage and presentation of information and data.

Some measure of growth of the medical literature is given by the increasing activity of the National Library of Medicine. The Library receives more than 18,500 different journals. Each year, the Medical Literature Analysis and Retrieval System (Medlars) indexes approximately 175,000 articles taken from 2,400 biomedical periodicals. About 45% of these are in languages other than English. According to Bulletin, National Library of Medicine, "Guide to Medical Services," published in 1967, the Medlars File contained approximately 486,000 citations on magnetic tape.

One measure of the national significance of biomedical research and medical practice and the supporting data activity is the dollar and manpower investment in this field. It has been estimated that some $2.3 billion or 6-10% of the total research and development in the United States was spent in 1967 for medical research. This is almost 10 times the amount of money spent for medical research during 1950. In the biosciences, industrial development has made a substantial impact upon drug development through the pharmaceutical industry. Industry is responsible for the support, through private funds, of about 25% of all national expenditures for Medical Research. In 1966, industry spent $500 million for medical research, and $560 million in 1967 (see Table II-F-1). Federal, state, and local governments contributed $1,380 million in 1966, and approximately $1,540 in 1967. The balance of the estimated 1966 and 1967 funds came from private support.

Federal support for medical research is a portion of many agency budgets. It is quite natural that the largest appropriation should come from the Department of Health, Education, and Welfare--an estimated $1,070.6 million for Fiscal Year 1967. Of this amount, 55.2% or $813.9 million will be appropriated to the National Institutes of Health. The Department of Defense has an estimated medical research budget of $114.3 million, the Atomic Energy Commission has $95.4 million, and the Department of Agriculture has $45.7 million. The National Aeronautics and Space Administration has a medical research budget of $79.9 million, and the Veterans Administration has estimated that it will have spent $46.9 million during 1967. Table II-F-2 lists the government agencies and the medical research during budgets for Fiscal Years 1966 and 1967.

TABLE II-F-1

NATIONAL EXPENDITURES FOR MEDICAL RESEARCH, 1966-1967
(In Millions)

SOURCE OF FUNDS	1966 est.	1967 est.
TOTAL:	$ 2,050	$ 2,275
Government	1,380	1,540
Federal	1,319	1,475
State and local	61	65
Industry	500	560
Private Support	170	175
Foundations and Health Agencies	95	98
Other Private Contributors	28	29
Endowment	19	19
Institutions' own funds	28	29

These statistics cover medical and health-related research only; activities such as research training and construction of research facilities are not included.

TABLE II-F-2

FEDERAL SUPPORT FOR MEDICAL RESEARCH, 1966-1967
(Dollars in Millions)

AGENCY	1966 Amt.	1966 %	1967 Amt. est.	1967 %
TOTAL	$1,318.7	100.0	$1,475.5	100.0
Atomic Energy Commission	89.6	6.8	95.4	6.5
Agriculture Department	44.7	3.4	45.7	3.1
Defense Department	119.2	9.0	114.3	7.7
Health, Education and Welfare Department	928.3	70.4	1,070.6	72.6
(National Institutes of Health)	(790.9)	(60.0)	(813.9)	(55.2)*
Federal Aviation Agency	2.5	0.2	2.7	0.2
National Aeronautics and Space Administration	75.4	5.7	79.9	5.4
National Science Foundation	13.8	1.0	14.4	1.0
Veterans Administration	40.7	3.1	46.9	3.2
Other	4.7	0.4	5.4	0.4

Covers only medical and health-related research; such activities as research training and construction are not included.

*Excludes mental health.

The largest segment of NIH money appropriated in 1967 went to cancer research ($176 million); next came heart research ($165 million); arthritis research ($136 million); neurology ($116 million); allergy research ($91 million); child health ($64 million); dental health ($28 million); and environmental health ($13 million).

The practical impact of the research sponsored by these several organizations can only be appreciated with some comprehension of the limitations of the community of practicing physicians. There are only 272,000 physicians in the United States to attend the health hazards of a population of 200,000,000.

This means that there is an average of one physician for every 7,350 persons, assuming that all medically trained persons were practicing medicine. But, of the total number of physicians, only 64,800 are General Practitioners or "family doctors." The balance of 207,200 specialize in narrow fields of medicine or are engaged in full-time teaching or research.

There are 11,721 Osteopaths practicing in the healing art. They are licensed to practice in all states, and many grant them the same privileges as those given to the M.D. Some colleges of Osteopathy have curricula comparable to that of a recognized medical school, or so it has been maintained. At least one state medical society (California) grants them admission, giving them equal status and benefits of full membership.

It follows that the 64,800 General Practitioners and 11,721 Osteopaths must attend to the majority of the sick in a total population of 200-millions and must simultaneously cope with the flood of data evolving from $2.3-billion of research activity.

While biomedical research serves as the primary source for most data of use in clinical practice, the use of data developed by other fields of scientific research allows the physician to improve diagnostic methods and enhance abilities to cure disease. Therefore, data efforts linking medical research with other scientific and technical activities are of great value in providing the medical profession with translatable techniques and applicable theories. Figure II-F-1 illustrates some of the modes of science and technology linkage.

2 Data Characteristics

There are two principal classes of data in the medical field, preclinical data and clinical data. The first of these provides the basis for understanding the effects of chemical and biological changes in the body. Information of this kind assists the physician and surgeon in the analytical and decision-making processes. These data are the basis for the basic medical sciences of physiology, anatomy, clinical and biological chemistry, pathology, microbiology, and neurology. They are largely descriptive in nature; for example, in the fields of pathology, microbiology, and anatomy, they may consist of specimens, prepared microscopic slides, or photographs.

Clinical data, which consist of sociological, scientific, and technical observations, assist the physician in making a diagnosis of the patient's disease and contribute toward the treatment regimen. Clinical chemistry reports, diagnostic photographs, tonometer readings, eye ground examinations, pulse and heart rates, respiration, temperature, psychological responses, and environmental conditions are examples of clinical data.

The relationship between these classes of clinical data and the data which result from preclinical activity is complex. It may be best described by indicating the broad spectrum of research activities which generate the preclinical data, the wide variety of specialty fields in medical practice which use these data, and, in addition the diverse types of clinical data. Figure II-F-1 illustrates their complex and multiple relationships.

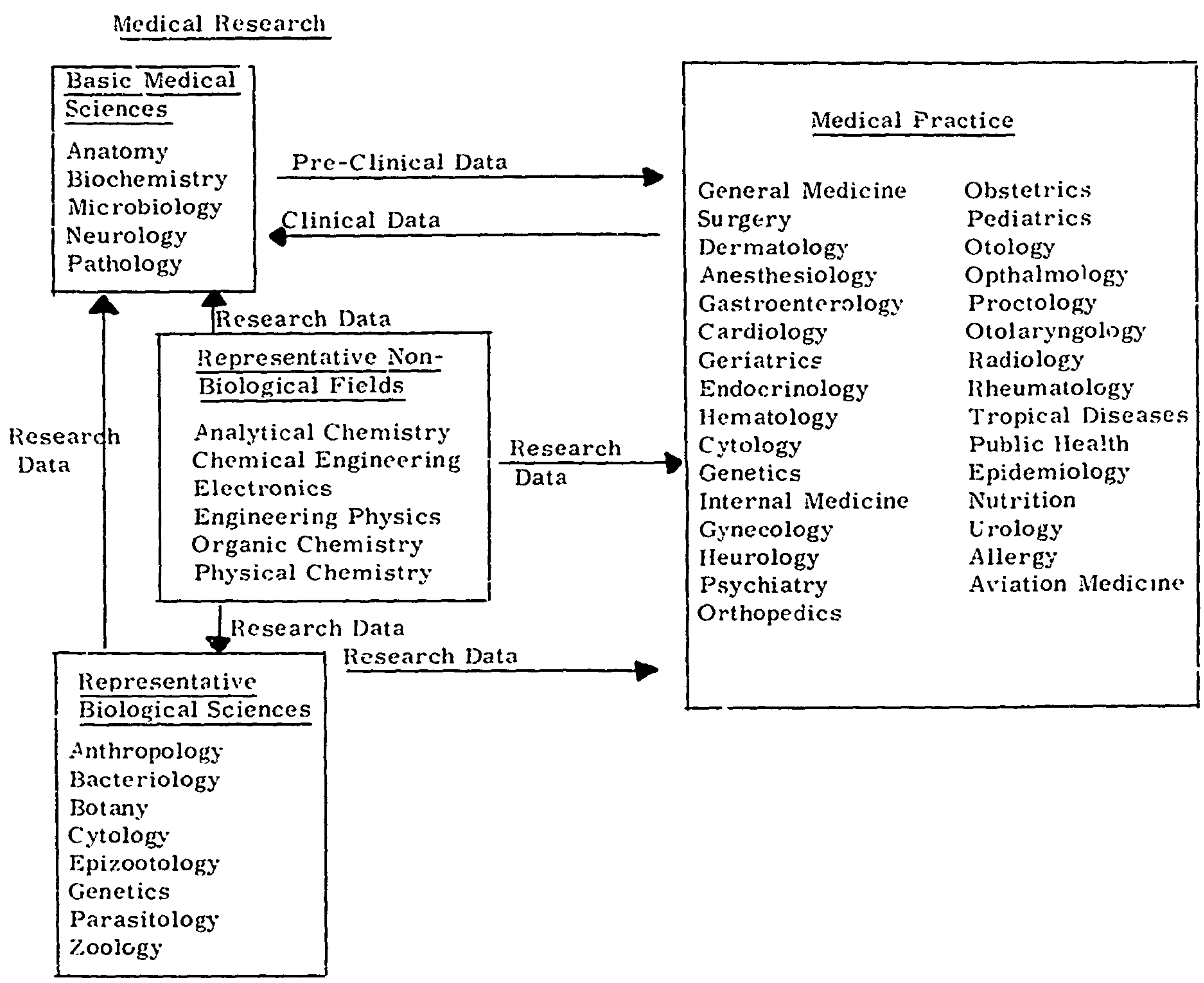


FIGURE II-F-1. FLOW OF DATA BETWEEN MEDICAL RESEARCH AND PRACTICE

Because biological systems are so complex, there are always many unknown variables, and for any given system at any time, behavior or response to treatment is seldom predictable with 100 percent assurance. Therefore, the utilization of both classes of medical data (clinical and preclinical) must be influenced by probability factors in much the same way as other types of scientific and technical data are used; for example, the validity of a treatment must be verified through testing in a large number of cases. The medical community makes full use of the probability factors which must be weighed and evaluated during every diagnostic procedure, and this aspect of medical practice has great bearing on the quality and often the form of the data.

Certain segments of medical data are numerical. Measurements of blood pressure, rates of blood flow, respiration rates, heart rate, and the results of clinical laboratory tests are expressed as numbers. However, numbers are often insufficient and do not thoroughly describe the circumstances. For example, while blood flow and heart rates may be expressed as numbers, the elasticity of the vessels is of extreme importance to the proper diagnosis. Similarly, while a blood cell count is significant and will be expressed as numbers, the size, shape, and color of the cells are also important and must be indicated with word descriptors. An example is crescent or sickle cell anemia, wherein the cells are shaped like a sickle; another example is aplastic anemia caused by a deficiency in the blood-cell-producing activities in the bone marrow.

The format for recording and presenting medical data includes photographs, graphic recordings, and statistical summaries. Photographs and 35 mm. slides play important roles in medical data activities. Excellent examples of these techniques are found in the American Registry of Pathology, a department of the Armed Forces Institute of Pathology. The 2x2-inch mounted transparancies contain pictures of clinical, gross, microscopic,

and X-ray material. These can be used in conjunction with microslides and a printed syllabus that accompanies each set. The American Registry of Pathology has produced a large number of clinico-pathologic conference sets in loose-leaf binders that contain all the clinical and pathologic facts necessary for a clinicopathologic conference. The sets cover various subjects and are loaned, without charge, to pathologists both in this country and abroad for a period of two weeks. Currently, this data collection contains 880 titles with 6,202 sets of slides.

An example of the use of graphic data is in electro-encephalography, which is the graphic recording of the electrical currents developed in the cortex by brain action. The recordings are tracings on long strips of paper. The height and the width between the lines are indicative of the patient's condition. Studies of this type of data have enabled the physicians to diagnose such diseases as epilepsy, chorea, and other conditions due to brain damage.

Electrocardiography is a method of making graphic records of the electric currents emanating from the heart muscle. It is a method for studying the action of the heart muscle. This type of data is necessary in order to know the organ's state of health. Another type of data useful to the cardiologist is the electro-cardiophonogram. This is an electrically activated recording of heart sounds which describe the functional activity of the heart.

An example of the use of statistical data is the field of epidemiology, which deals with the relationships of the various factors which determine the frequencies and distributions of an infectious process, a disease, or a physiological state in a human community. Obviously, the results of any epidemiological study are statistical. An excellent example of such statistical data is to be found in the National Disease and Therapeutic Index. This is a continuous survey and study service to determine the types of diseases found in various parts of the country, the incidence of the conditions, frequency, and types of treatment. Answers to these, as well as other questions, are given in numbers and percentiles.

Proprietary Considerations. Biomedical information and data, as is the custom among most other sciences, are freely exchanged among the professionals. The exchanges may take place in an informal manner at seminars, meetings, conventions, or by means of letters and telephone conversations. The formal data are published in journals, textbooks, reports and other types of printed media and, as is occurring at an increased rate, as computer print-outs.

While a good deal of the published material is copyrighted, permission to quote and reprint is customarily granted freely. Appropriate credit must be given to the author and publisher of every quotation. Obviously, the data generated by the government agencies are in the public domain, with the possible exception of material which is classified for reasons of national security.

A most important consideration in the handling of all medical data is the protection of the patients' right to privacy. Data concerning individual patients are not discussed or passed along without the individual's permission. When discussions take place or a paper is written, it is the customary practice to maintain the patients' anonymity. In the future, if clinical data networks are established, this principle should be followed.

2. Medical Data Flow

Advances in medicine have been made along three main avenues, parallel with concomitant generation and use of medical data. The methods used are: (1) clinical or bedside observations and the treatment of sick individuals based on "patient data", (2) laboratory experiments and observations, and (3) statistical measurements and analyses of the characteristic patterns of human health and disease.

Medical history is replete with examples of accurate clinical observations or patient data which resulted in subsequent successful treatment of injured and diseased persons. A superb example of accurate, keen observation applied toward the elimination of smallpox was Edward Jenner's (1749-1823) scrutiny of the girls who milked cows infected with cowpox and who were immune to smallpox. To this day, his collecting procedures and his laboratory methods for the preparation of smallpox vaccine are but slightly changed. Jenner's observations, laboratory experiments, and his subsequent communications and publications represent the classical use and flow of medical data toward the prevention of disease.

The flow of "patient data" from medical practitioners is largely on an informal basis, although there is one formalized effort, The National Disease and Therapeutic Index (NDTI), a private enterprise, which compiles data from private practitioners. Financial support comes from about 45 ethical pharmaceutical manufacturers who have a need to know all about disease trends, changes in disease patterns, and the disease treatment requirements. The NDTI reports are used by the pharmaceutical companies as an aid toward directing their research and development programs. Activities began in 1956, and in 1960, they began placing the data on computer tapes. They now have over 2 million patients' visits on the tapes. The data can be analyzed in many different ways and will provide information tailored to fit the subscribers' needs.

Patient data are continuously collected from participating private practitioners on two days of their practice at four different intervals of the year for a total of eight days of all the patient visits. This includes patients seen at home, in the hospital, in a nursing home, contacted by telephone, or in an office visit. The regional panel of 1, 500 physicians is continuously changed so that each quarterly report statistically represents a very large proportion of the medical population. When the data arrive at headquarters, they are coded and fed into the computers.

NDTI has developed formulae for statistical evaluation of the data base. For example, one region may have 4,000 internists who, due to the rotation of reports, will send data for every day of the week, Monday through and including Sunday, 31 days in the month. NDTI might use 15 internists who will report on two days' practice in the same region. The projection then would work out in a manner similar to this: 31 days x 4,000 practitioners divided by 15 x 2 would give the factor, which in this instance would be 4,130; that factor is then multiplied by the number of mentions of a disease to give a useful statistic. For example, should obesity be mentioned ten times by this sampling of physicians, then 10 x 4,130 would correspond to 41,300 visits by obese patients to internists in this particular region. Such national estimates are found by using the same basis for each region and adding all the regional sums. The standard error in these calculations is usually about 4%.

Such statistical data indicate trends for various diseases, regional disease patterns, seasonal disease patterns, unique characteristics of diseases, patterns of drug usage, areas where better treatment is required, disease difference by patients' sex and age, and treatment data on the Medicare patient. The data are collected and reported on 22 variables at the time of the patient visit.

While the service is essentially for the benefit of the organizations who financially support the effort, any medical practitioner who desires to obtain data may do so without charge. Physicians request regional profiles on various diseases, profiles on patient characteristics, disease trends, and average patient loads both within their region and in other regions.

Another major generator of patient data is the hospital. Practically all hospitals of any reasonable size use automatic data processing to handle their accounting activities. Information taken from these data tapes can be used to analyze such items as the length of residence in the hospital, duration of disease, types of surgery performed, kinds of disease treated, and comparisons with epidemiological statistics from larger segments of the population, as well as at the national level.

Large hospitals (those having 200 beds or more) with teaching facilities and outpatient departments use the ADP facilities to assist in the clinical research studies conducted by the various specialists and department chiefs. Research activities are carried on in all areas of clinical medicine. The data evolving from these many efforts usually result in papers published in one of the many medical journals. While the data stored on the tapes may again be used, they do not represent material of national significance unless the study has resulted in some new, useful information, or a scientific discovery has taken place.

Statistical data are also generated by several organizations to determine the value of a therapeutic regimen, the success of a surgical procedure, the values of a laboratory experiment, and the behavior of populations in health and disease. An excellent example of the flow of statistical data is the Commission on Professional and Hospital Activities, a non-profit organization located in Ann Arbor, Michigan. It has the official sponsorship of the American College of Physicians, the American College of Surgeons, and the American Hospital Association.

The Commission gathers, classifies, and analyzes statistical data from 1,050 hospitals located in 46 states, the District of Columbia, Hawaii, Puerto Rico, Australia, nine provinces in the Yukon Territory, and five hospitals belonging to the Arabian American Oil Company. These hospitals provide the Commission with about 24% of all the short-term general hospital discharges in the United States and Canada. There is an annual volume of 8,200,000 hospitalizations on which they obtain abstracts of the case history. They receive identifying information on each patient: date of admission and discharge, age, sex, doctors who attended the patient, data of surgery, anesthesia, tissue findings, the birthweight and sex (if it is a newborn), San Francisco systematized code for tumors, admission findings (e.g., temperature, total white blood cells, hemoglobin, hematocrit, blood pressure, urinary findings such as glucose and albumen), the highest temperature the patient had during hospitalization, and the blood sugar levels. There are approximately 60 investigating procedures and drug data on 15 general classes of medications. All the data are classified according to the Armed Forces Standard Systemized Code for Pathology and the International Classification of Diseases Adapted.

The Commission uses 265 employees to collect and maintain all these data. These people develop a variety of reports of value to the hospital medical staff and to the management. For example, they supply mortality statistics, and numbers of patients admitted with specific types of diseases (disease indexes); total number of surgical pro cedures, and kinds of surgical procedures; types of therapy employed, where therapy was used, and results of therapeutic regimen; number of patients in each decade of age, and average stay within each decade, and frequency of incidental or secondary diagnosis. The monthly reports are also recapped at a 6-month interval, giving rates, averages, and percentages.

Statistics of this nature are extremely useful to the various medical staff committees. They report the incidence of disease, length of time required for treatment, quality of the therapy or the surgery, and such items as the number of transfusions, tissue grafts, and bone grafts. These epidemiological data flow from the patient diagnosis not on the chart, to the treatment regimen or surgical procedure, to notes on the convalescent period, and then to the data acquired before and after discharge from the supervision of the physician or surgeon.

Another example of a statistical data generator is the American Dental Association, which is now evolving a dental medicine data center. Included in the main body of dental data are two classes: dental materials data and clinical evaluations. Information and data on dental materials are gathered from many sources--publications, manufacturers' literature, and the dental materials research section of the National Bureau of Standards, but wholly supported by the ADA. The Dental Materials Handbooks are published by the ADA. Oral diseases also receive a great deal of attention, but progress is painstakingly slow and adequate results are difficult to obtain. The results of dental clinical studies may be found in the various dental journals.

Another field in which statistical biomedical data are being generated is that of veterinary medicine. There are great differences between the systems of practice conducted by the veterinarian and the physician. There are, of course, the obvious differences between the human being and the animals, but there are also great differences in the types of diseases which infect different animals. Furthermore, the veterinarians have few large hospitals or clinics. The exceptions are those maintained for teaching purposes by the veterinary medical colleges. The veterinarian is essentially an individual, lone practitione with little contact among other veterinarians, whereas the

physician sees, converses, and consults with his colleagues almost every day. The physician contributes a certain amount of his time and skills to the free clinics and will engage in clinical research as an exercise in self-education and as an avocation. The veterinarian cannot contribute any of his time and talents to non-existent free clinics or general hospitals. With the exception of some ASPCA facilities, there just are no free animal clinics of national significance. Practically all veterinarians are in private practice. Many are employed by the Department of Agriculture as animal and meat inspectors, in NASA to attend the experimental primate colonies, in the pharmaceutical firms to assist in the development of veterinary pharmaceuticals, by the Army Veterinary Medical Corps, and by the veterinary colleges of medicine as full-time instructors and professors. Many of these groups make research contributions.

There are two formalized data efforts in this area. One is located in the San Diego Zoo and is called the Morbidity and Mortality Data Center. It collects epizootiological data on zoo animals. The other one, the Veterinary Medical Data Center, is under the aegis of the National Cancer Institute and is located in Michigan State University's College of Veterinary Medicine. It collects and analyzes epizootiological data with the cooperation of the veterinary medical schools at the Universities of Minnesota, Missouri, California at Davis, Pennsylvania, and Duke.

Veterinary statistical data are disseminated in much the same manner as in the other biosciences. The effort begins in one of the laboratories, progresses to clinical studies, and then to publication. The official journal, and probably the best one in the United States, is the Journal of the American Veterinary Medical Association. There, one is likely to find reports on the best research.

Influence of ADP on Biomedical Data Flow. While the preponderance of biomedical data flow via the printed media (textbooks, references, reports, and journals), there is an increasing trend toward the use of electronic data processing equipment to alleviate the problems of clinical decision-making, which involves an enormous number of variables. For any given clinical diagnosis and prognosis, there may be as many as 100,000 observable findings, 10,000 known diseases, and 100,000 available treatments. Acquisition, storage, and evaluation of information needed by the practicing physician to assist in clinical decisions through use of computers have begun to find some application. While computer systems were originally used in hospitals to perform accounting and administrative functions, the present trend is to use these machines as a research aid and to assist in patient care.

During a Workshop on Medical Data held during the A.M.A. meeting in Atlantic City, N.J., June 19, 1967, Dr. Jordan J. Baruch of the General Electric Company's Medical Division made these relevant remarks about medical data systems:

> "Much of the excitement in the early 1960's in the computer information systems applications to medicine centered around artificial intelligence and complete automation... In medicine the attention was focused on automatic patient monitoring and complex physiological control systems. The significant meetings of the early 1960's were concerned with automatic diagnosis of disease, given a set of symptoms or the automatic recognition of complex physiological conditions given as the output from a set of sensors; i.e., high-blood pressure sensors attached directly to a computer. Then came the machine interpretation and control of

medical data; the use of machines led to a reassessment and growing appreciation of the skills of people. In medicine, for example, remote monitoring, except for very special cases, quickly gave way to the return of the nurse. She was required to perform complex pattern inter-correlations necessary for a useful assessment of the patient's condition. Machine diagnosis was gradually transformed to machine-aided diagnosis as the complexity of the pattern became more apparent and as symptom validation and context relations among symptoms were recognized as important. . . . The reinclusion of man in the information processing loop has by and large taken the form of an increased interest in iterative or interactive systems. On-line systems have been implemented and are being tried in order to find out how best to integrate the joint efforts of man and machine.

"One would like to believe that in the humane art of medicine, man has reappeared on the scene because of some innate humanist attribute of the problem. Alas, such is probably not the case. At the present state-of-the-art and under present economic constraints, people are simply cheaper to use than machines for many complex or heuristic processing operations. Even a well-paid radiologist costs less to operate and reprogram with changes in medical knowledge than any pattern-recognizing film reader. Even if the radiologist were not performing a far broader medical service than simple film interpretation, simple economic comparison has cut down the level of interest in automated film readers.

"There are areas of medical data handling where the machine can act as an adjunct to the human in tasks that intelligent humans seldom do especially well. The areas of sorting, filing, indexing, searching, and particularly of being alert for low probability occurrences are the kind of performances that hardware can do well and that intelligent people do poorly. There are a number of centers which report clinical chemistry data. In these places the computer is used in part as an aid to communcation, in part as a neat and easily maintained file cabinet, and in part as an automatic alarm generator. Test values outside of a clinically acceptable range raise general alarms and do so unfailingly regardless of the low probability of such an alarm's actually occurring.

"Great values are placed on machine-stored data systems since such records are available for machine searching, manipulation, and retrieval, and this availability for intensive analysis is valuable."

The objective of automated medical data systems is to quickly and easily provide the medical practitioner and researcher with ready access to the most recent medical knowledge and to develop diagnostic parameters which will aid his decision-making abilities. Practitioners may also be relieved of activities that can be delegated to para-medical personnel and technicians. The use of a medical data system increases the physician's effectiveness by permitting him to concentrate on tasks requiring his special insight, skill, knowledge, and understanding. These same data systems may also be used as educational tools and to assist in clinical research. Although automatic data systems will increasingly exert important influences upon medical practice, this influence would come mainly in the area of the physician's present strength. That is, it would buttress his knowledge and augment his repertoire of technical skills, but it would not change the basic nature of his practice or his approach to his patients.

. Automated clinical test procedures conducted by means of an auto-analyzer connected to a computer may routinely perform a battery of ten or more chemical screening tests on human blood at the cost of doing two or three such tests in the usual manner. The use of low-cost automatic multiple screening tests uncovers a significant number of unsuspected abnormalities. This would also occur were these tests to be done on a one-at-a-time basis. However, the demands on the patients' time and funds are so great that these are done only under very special circumstances. By means of the rapid automated procedures, the physician is able to see and treat a greater number of patients in the presymptomatic or asymptomatic phase of the disease. Thus, the serious and often painful symptomatic phase of the disease is frequently prevented. Automated multiphasic screening systems supplement the physician's diagnostic skill in the interpretation of symptomatic disease with information that is useful towards the maintenance and the early recognition of disease.

Automated medical data systems are used in three areas:

- Biomedical research,
- Diagnosis and treatment of patients, and
- Administrative-accounting-population statistics, integrating administrative functions in hospitals and clinics.

At present, automated medical data systems are widely used in biomedical research. This is largely due to the machine ability to speedily and precisely analyze huge quantities of data and to the development of new mathematic techniques and concepts. The most obvious use of computers in research is in the manipulation of large volumes of information and complex interrelationships. An example of such an application is the long-term study of heart disease at UCLA, in which 1,000 items of data are collected for

each person in the study. Another is the study of Patterns of Infectious Drug Resistance in one General Hospital. These tests were conducted on five different antibiotics and various combinations of these under a variety of conditions. Without the use of computers, the manual effort involved in compiling the data might have eliminated the possibility of research.

The future role of automatic data processing and formal data efforts in the management of both clinical and preclinical data will be to a large degree influenced by the difficulty of separating data generation, communication and usage functions. The following discussion identifies presently discrete data generation, dissemination and usage activities.

Data Generation - Hospitals and Schools. Proficient medical teaching requires close associations with hospital and out-patient clinics. Therefore, all medical schools are associated with a large hospital. The teaching staffs of the schools, as well as hospital staffs, are intensely interested in research and, of course, in acquiring the data generated from other areas in the sciences. There are close communicative associations between the research physician and the biomedical scientist or engineer, and the close association among these groups eases the flow of basic scientific findings toward clinical application. Furthermore, since a very large proportion of biomedical research is conducted by physicians with dominant commitments to patient care, the clinical use of research findings closely follows the disclosure of new, useful data.

The Veterans Administration operates a huge group of medical facilities, and is representative of hospital and medical school cooperation in teaching, research, and patient care. It is, therefore, typical of the communication situation in teaching hospitals' activities and the research conducted by those institutions. The Veterans Administration Hospitals system is an

enormous resource for biomedical research and health personnel training. It operates 165 hospitals providing more than 120,000 beds; 211 outpatient clinics that handle more than 6 million outpatient visits a year; 16 domiciliaries that house about 14,000 members, and more than 30 nursing-home facilities. The patient-care facilities are, in the aggregate, slightly larger than the total non-Federal major hospital affiliates of the medical schools.

The VA research programs are essentially disease-oriented, and they involve about 7,000 research projects. These are conducted in 146 hospitals and 12 out-patient clinics. About 3,000 physicians and 1,000 scientists are involved. During the past year, these programs have yielded more than 3,000 publications in professional and scientific journals as well as thousands of verbal presentations, exhibits, and other forms of communication. More than 85% of the total effort is clearly identifiable as applied clinical research.

The VA employs about 6% of the nation's medical manpower, including more than 2,000 physicians who hold active academic appointments in medical schools and universities. Presently, 88 of the VA hospitals are affiliated with 74 medical schools. They are also affiliated with 32 of the Nation's 47 dental schools, all of the 56 accredited schools of social work, the 58 universities approved for graduate training in clinical and counseling psychology, the 145 basic nursing progams, and 127 schools that provide clinical training in physical medicine and rehabilitation.

The VA maintains data and information flow to scientists and engineers, internally and externally, primarily by means of its publishing activities in scientific and professional journals. During the fiscal year of 1966, the VA program evolved 3,417 published articles. Many papers were presented before lay

groups and scientific meetings, where informal exchange between investigators is standard practice. The VA holds medical research conferences which provide intramural exchange of information, and some exposure to other information is obtained through the invitation of prominent non-VA scientists. The 2,000 VA physicians who hold academic appointments contribute biomedical information as teachers.

At the VA, an Office of Scientific Communications was established and is active in developing more extensive dissemination of biomedical information. This Office channels information concerning VA and VA-sponsored research to the Science Information Exchange of the Smithsonian Institution. In addition, the Office writes and disseminates brochures and pamphlets about aspects of research. It publishes a bimonthly Research and Education Newsletter, and furnishes weekly highlights about research and cooperates with the Office of Information Service in publicizing newsworthy projects. Circulars and medical bulletins also serve to distribute needed information to field stations. A Termatrex data and information storage and retrieval system is under development for technical information. An Automated Research Information System (ARIS) is also under development, and when operating, will quickly furnish needed information about the research programs.

Data Generation - Clinics. The advent of automatic data processing equipment has introduced a new epoch into clinical medicine. This equipment can remove from the physicians' direct supervision many tests by permitting technicians to perform routine tasks and analyses. Thus, the physicians' time is better and more profitably utilized for diagnostic and judgmental functions, and test data are automatically stored and retrieved as required for evaluation. The combination

of automated analytical chemical analysis and computers has already resulted in medical systems which perform repetitive procedures with rapidity and accuracy impossible of attainment by humans. These systems permit the mass production of more data on more people and also more data on each individual patient. Medical data systems of this type increase the feasibility of more explicit evaluations of the patients' medical requirements and provide a more accurate appraisal of their biophysico-chemical values.

The Kaiser Foundation of San Francisco is now operating an Automated Multiphasic Screening System as part of a diagnostic program called "The Multiphasic Health Checkup." The system is capable of handling about 30,000 patients per year with a minimal number of physicians. The testing period requires two or three hours. The patient receives a battery of some 13 tests, such as an electrocardiogram, X-ray, pulse and blood pressure, visual acuity, respiration, and retinal photography. Eight blood analyses are performed within 12 minutes by an automatic chemical analyzer. The patient is given a urinanalysis, red and white blood cell counts, blood grouping, and a serological test for lues.

The foregoing tests are conducted by technicians. Then the patient is given a complete physical examination by an internist. All the data from the various tests and examinations are fed into the computer and are printed out for the physician at the last position where he evaluates the findings and may recommend any additionally required tests such as a gynecological examination, usually a cervical smear for cancer detection, a cystoscopy or a sigmoidoscopy.

The creator and director of the Permanente Medical Group of the Kaiser Foundation, Morris F. Collen, M.D. has this to say:

> "The physician who is skilled in diagnostic evaluation of the complex physiological systems of the human body is probably the best-trained systems analyst in present society. Once cybernated systems applicable to medical science become generally available, the practicing physician will not only readily accept and adjust to them, but will soon demand these services for his patients. In the future, it is likely that every community of 100,000 or more will have affiliated with one of its larger hospitals an automated multi-test laboratory, which will be available for admission examinations and preoperative examinations for the hospital patients, and which will be utilized for office patients for periodic health examinations, general health examinations for special purposes (industrial, insurance, etc.), early sickness consultations and diagnostic surveys. These multi-test laboratories will undoubtedly be affiliated with a regional computer center which will provide data processing services through connecting telephone lines."

In another paper, Dr. Collen made these remarks:

> "If periodic health examinations are to be provided to large numbers of people at a reasonable cost, the use of an automated multitest laboratory has several advantages. (1) improved efficiency of service to patients through close integration of many test procedures; (2) improved efficiency for physicians by providing, with the first office visit, a large amount of information about their patients; (3) improved quality control with automated equipment; (4) improved economy by providing at least four times as many tests for the same cost and at a greater speed; (5) earliest possible detection of a wider range and greater number of unsuspected diseases among apparently healthy people; the concept of health evaluation in addition to disease detection becomes possible by providing the physician with a more comprehensive profile of the individual patient's physiological state; and (6) computer data processing and computational capabilities permit multivariable epidemiologic research heretofore not possible."

The need for automated clinical testing, analytic and data systems to diagnose disease and to act as an aid in the prevention of disease by revealing pre-clinical symptoms is recognized by qualified men of medicine. Dr. Robert M. Zollinger, Professor and Chairman of the Department of Surger, Ohio State University, said, at a recent White House Conference on Health:

> "No physician can today or in the foreseeable future have the time to take care of his patients, and he must depend upon auxiliary help. I foresee that, by special training now proposed for the physician in family practice, he will serve more and more as triage officer by directing his problem patients to special centers for definitive treatment."

Surgeon General William Stewart said at the same conference:

> "Year by year, our top professional personnel are being trained to perform still more complex tasks. How long can each profession afford to hang onto its simpler functions - the routine filling of a tooth, for example, or the several easily automated steps in a medical examination? How can we train the physician or dentist to make full use of the skills available in other people, freeing himself to perform only those duties for which he is uniquely qualified?"

Answers to these questions may be in the automated clinical testing laboratories and in training programs for para-medical personnel.

Data Generation - Private Institutions. The private research institution may be best exemplified by the Worchester Foundation for Experimental Biology of Shrewsbury, Mass. It is an entirely independent, non-profit research institute. It deals with steroid chemistry and its application to medicine, neuroendocrinology and mammalian reproduction and population control via collaborative working arrangements with half a dozen private and State hospitals in the area. While the Worchester Foundation is concerned with basic science, their collaborative relations with hospitals give them opportunities to communicate and apply basic discoveries. An example of the application of a basic discovery toward practical application of the findings was the discovery that synthetic steroids will block ovulation in mammals. This led to the practical use of the now widely known and used birth control tablets.

The professional staff consists of organic chemists, biochemists, physiologists and internists. Thus, the research results and data are published in a variety of scientific journals as well as in their own bulletins.

Data Generation - Federal Laboratories. The Federal Government operates several laboratories that play a primary role in biomedical data generation. An example is the laboratory operated by the Atomic Energy Commission's Division of Biology and Medicine. It has the responsibility of studying the dose-effect relationships between radiation and living things. In the matter of health and safety, the Commission has the responsibility to assure that daily association with nuclear energy will be carried out in a safe and responsible manner. This implies an ever-growing body of information on the mechanisms of interactions of radiations with tissues, cells and molecules. The Division of Biology and Medicine has the responsibility of developing these data through its basic research program, as well as exploiting nuclear energy and its by-products in medicine and biological research.

Research is conducted in 14 areas, and data are generated and published from each effort. The divisions of effort are as follows:

- Molecular & cellular level studies;
- Radiation genetics;
- Somatic effects--general;
- Toxicity of radioelements;
- Environmental radiation studies;
- Radiological physics;
- Health physics
- Radiation instruments;
- Combating detrimental effects of radiation;
- Chemical toxicity;
- Nulcear energy, civil effect;
- Atmospheric radioactivity and fallout;
- Cancer research; and
- Applications research.

Among the biomedical developments which have evolved from this research effort are electronic devices for segregating cells in pathological investigations, teletherapy units to supplant high-voltage X-ray machines, and scintillation cameras for diagnostic scanning of human organs.

Data and information emanating from these and other AEC research efforts are published in scientific journals and presented at national scientific meetings and conferences sponsored by biomedical societies and agencies supporting biomedical research. Additional data are disseminated in the AEC's report Fundamental Nuclear Energy Research, available without charge from the Superintendent of Documents. Nuclear Science Abstracts, also available from the same agency, provide a key to certain sources of substantive data evolving from AEC research programs relevant to the biomedical field.

Data Flow Intermediaries. Collection networks, data centers, document depositories, the published literature and informal sources are the principal intermediaries in the flow of biomedical data. Of the five classes of intermediaries, the last three are by far the most significant. Only one data network and twelve data centers were identified in this study; whereas, the influence of depositories, published literature, and informal sources was found to be so great that it was difficult to adequately assess them.

The data network which was identified was the Veterinary Medical Data Program at Michigan State University, College of Veterinary Medicine. It collects data pertaining to naturally occurring diseases of domestic animals from eight United States colleges of veterinary medicine and one located in Canada. Funding is provided by the National Cancer Institute at the National Institutes of Health, which are under the Public Health Service. The premise for its operation is the paucity of animal disease data in other areas, such as cancer research. According to Dr. James A. Peters of the Epizootiology Section of the Natural Cancer Institute's Epidemiology Branch,

> "Additional justification was based on: (1) recent impetus for research on chronic degenerative as opposed to acute infectious diseases; (2) an emerging concept of medicine as a single discipline; (3) increased interest in, and expansion of, comparative medical research; and (4) knowledge that distribution and occurrence of disease first become manifest by the mechanism of collection, storage, manipulation, and recall of clinical disease data."

In addition to creating the impetus for the Veterinary Medical Data Program, these factors have led to the development of the several data centers in the medical field.

- The Automated Hospital Information System (AHIS), operated by the Veterans Administration, supports ongoing hospital operations related to patient care.
- The Adverse Reaction System and Drug Application System, operated by the Food and Drug Administration, provides an early warning system for detecting unknown adverse effects of drugs.
- The National Center for Health Statistics, operated by the United States Public Health Service, generates nationally significant health statistics.
- The National Disease & Therapeutic Index (NDTI) is a private enterprise which gathers data on the distribution of diseases in the United States and the types of therapy used as treatment.
- The Hospital Purchasing File contains information on all products, equipment, instruments, supplies and materials used in hospitals and nursing homes.
- The National Formulary, produced by the American Pharmaceutical Association, is an official directory of mixtures used in medicine.
- The United States Pharmacopoeia, provided by the United States Pharmacopoeia Committee, is an official and legal directory of pure compounds used in medicine. It contains the monographs of purity and the official testing and analytical methods for the determination of the parameters.

- The Mapping of Disease Project of the National Institutes of Health is under auspices of the United Nations, World Health Organization (WHO). It develops the world's geographic distribution of diseases.

- The Registry of Tissue Reactions to Drugs of the Armed Forces Institute of Pathology contains data on the reaction of tissues and organs to drugs.

- The American Registry of Pathology of the Armed Forces Institute of Pathology contains 2"x2" slides in color and black and white, photographs and descriptions of macroscopic and microscopic pathology.

- The Armed Services Pest Control Board, provided by the Research and Development Command at Walter Reed Army Medical Center, contains economic, biologic and medical information pertaining to the control of pests (arthropods, mammals, leeches, birds, reptiles, molluscs, plants, fungi) of military importance. This includes organisms related to (a) disease agents, vectors and reservoirs, (b) harmful and venomous effects, (c) damage to bionomics, taxonomy, control of pesticide toxicology, chemotherapy and limited basic research categories such as physiology. The Armed Services Pest Control Board supplies quarterly bibliographic citations on current literature; answers specific inquiries; data and information compilations by countries. The Board also publishes <u>Arthropods of Medical Importance-Series ES32</u>, available through the Earth Sciences Division, U.S. Army Natick Laboratories.

- The Commission on Professional and Hospital Activities is a private non-profit enterprise. It collects and distributes, to subscribers, data on ongoing hospital activities related to patient care. Hospital subscribers are located in the U.S., Canada and the Near East.

While these formal data efforts are playing an increasingly important role in the dissemination of biomedical data, the depositories which provide data-containing documents, publishers of biomedical literature, and informal sources continue to be the primary intermediaries for data flow. In effect, all biomedical libraries are also data-document depositories. The largest one is the National Library of Medicine in Bethesda, Maryland. The Library receives more than 18,500 scientific periodicals and publications. As mentioned earlier, the Medical Literature Analysis and Retrieval System (Medlars) annually indexes 175,000 articles taken from 2,400 biomedical journals. As of January 1st, 1967, the Medlars File contained approximately 486,000 citations. This enormous store of biomedical bibliographies is computerized and printouts are available as special, specific references upon request.

The John Crerar Library in Chicago, Ill. is similar in many ways to the NLM, but it is not as large, nor does it have electronic data processing equipment. It does house, however, a very large and valuable active collection of periodicals, texts, and references. The Technical Advisory Center for Lawyers in Philadelphia, Pa.; the Clearinghouse for Federal Scientific and Technical Information in Springfield, Va.; the Biosciences Information Services in Philadelphia, Pa.; and the Technical Information Division at Edgewood Arsenal, Md. are other examples of document abstracting and depository organizations that aid in disseminating data contained in documents. While it is true that the data content of the documents is not indexed in any of these organizations, the data are available through a traditional two-step search involving subject search via reading abstract printouts, and followed by data search in the pertinent documents.

Inherent in the use of document depositories as a data flow intermediary, is a considerable lag time from data generation to use. The time required for document generation, document dissemination, abstracting and input into depositories may be as much as five years. One shortcut is direct acquisition of published works from publishers. The problem is that the text and reference books, manuals, and reports number in the thousands each year and no single data user can acquire the documents he needs. It is not within the scope of this study to characterize all of the major publishing activities that support the biomedical field. However, the following list of principal book publishers does indicate the enormous number of book sources just in the United States.

D. Published Works

The number of texts, manuals, and reference books in the biomedical sciences ranges into the hundreds, perhaps thousands. As this report is essentially a census of data activities and not an index of publications, we shall only list the leading publishers.

Academic Press
Blakiston & Co.
Elsevier
Harper & Row Publishers, Inc.
Interscience Publishers, Inc.
J.B. Lippincott Co.
Little, Brown & Co.
C.V. Mosby Co.
McGraw-Hill Book Publishing Co.
Pergamon Press
Chas. A. Thomas Co.
John Wiley & Sons, Inc.
W.B. Saunders Co.
Williams & Wilkins Co.

As mentioned earlier, one of the primary sources of biomedical data is informal communication. According to James D. Watson, who was awarded a Nobel prize for his double-helical model of the DNA molecular structure, only part of the data he and his collaborators used came from formal channels of publication. In his book, "The Double Helix", he points out that:

> "Some of the salient information traveled on grapevines of personal relations giving fact and rumor about who was doing what that might be pertinent to their own work. Here, too, kinship ties could occasionally be utilized to advantage."

Among the more important channels for informal data communication are activities in biomedical and related societies. These societies conduct seminars, meetings and conventions devoted to the specific interests of their members. Such gatherings of scientists are most important in providing a forum for the presentation of papers, and opportunities to meet on a person-to-person basis and discuss topics of mutual interest. Informal discussions often result in lifelong professional associations wherein the individuals exchange ideas and problems over the telephone or by an exchange of letters. Frequently a group having a specific and highly specialized interest will band together to exchange ideas and information. Specialists in such narrowly defined areas are usually small in number and therefore do not form societies. They do, however, continue to exchange data and information on an informal basis, . . . sometimes for many years.

Many of the papers presented at meetings are never published in the scientific journals. Many such dissertations become available as unedited preprints, abstracts, or copies of the authors' talk. Conversations with authors often result in long term informal scientific exchanges.

An informal mechanism of great importance to scientists is the exchange of offprints and reprints. Offprints are additional printings of articles appearing in scientific journals. They are produced from the same plates as those used in printing the journal and they are printed during the production of the journal. Reprints are reproductions of original articles and are produced after the journal has been published and distributed.

Many of the public and private scientific research institutions produce bulletins and small magazines describing their research activities. Many medical schools and hospitals have monthly bulletins that are very valuable. The Bulletin of the Johns Hopkins Medical School and the Bulletin of the Yale Medical School are good enough to attract subscribers who pay a fee to receive these magazines.

The pharmaceutical companies publish and distribute, to the medical professions and to the hospitals, current awareness bulletins and magazines. Some of these publications feature special articles on topics of current interest and value while others contain abstracts of the most valuable articles in the current literature. As these efforts are in the nature of an essentially altruistic service to the medical profession, they contain data and information of value to the recipients, and are frequently devoid of product promotion. Informal distribution of clipped articles from these publications is an important mode of biomedical data flow.

There are many thousands of biomedical motion pictures available for viewing by the lay-public, students, scientists and medical practitioners. The Public Health Service, Audio-visual Facility located in Atlanta, Georgia, supplies, upon request and without charge, motion pictures on subjects such as microtechniques in serology, anesthesia, asthma, blood flow, environmental health, pedodontics, malaria, common cold, arthritis, diabetes, human chromosomes, primates, human growth, mental retardation, aging, smoking and health, nursing and home care, and epilepsy. They have a series of films on Heart Disease, Cancer and Stroke.

The American Medical Association also has a valuable film library. It covers all medical subjects of interest to physicians, surgeons and the para-medical groups. Loan copies are obtainable, without charge, and an extensive catalogue may be had for the asking. Many of the surgical supply manufacturers and the pharmaceutical houses have produced color and sound motion pictures depicting the use of sutures, instruments or drugs. While these are product-oriented, they do contain a great deal of valuable data and information, all depicted in a clear, graphic manner which may clarify some very complex subjects. All may be obtained from the producers without charge.

Biomedical Data Users. The multiple interrelationships between medical researchers and practicing physicians and the clinical and preclinical data make the job of characterizing biomedical data users nearly impossible. An analysis of the distribution of physicians' roles in 31 medical specialties shows that most of them are involved in various combinations of private practice, hospital residency, hospital staff work, teaching, and/or research. Some medical fields are so narrow that most physicians engaged therein work in other fields (i. e., aviation medicine, pediatric allergy, pulmonary disease) and their data requirements are accordingly distributed over several fields. Table II-F-3 summarizes the principal roles of biomedical data users in 31 selected fields of medicine. It indicates the spread of activities of each type of user in practice, residency, staff work, teacning and research, and therefore, the spread of data requirements.

Obviously, the data requirements for research and practice differ significantly, and the time demands placed on the practicing physician involved in other activities precludes the possibility for extensive data search. This is one of the problems highlighted in the following section.

4. Principal Problems in Biomedical Data Management.

A survey of data activities and associated problems in 87 active medical schools in the United States indicated the following problems and recommendations from 44 institutions:

- There is a lack of scientific personnel of the calibre, imagination and foresight to effectively utilize capabilities of automatic data processing systems in biomedical areas. Therefore, the universities must develop training programs slanted toward computer science and mathematical modeling in the life sciences. Support for this type of training program is immediately required.

- There is a requirement for standard nomenclature and coding systems to be used in medical data systems. Much of the data is descriptive and does not lend itself easily to coding for the computer, and these data must be substantially coded.

- There is a need for research methods for training the physicians to modify their language to make it more easily acceptable for coding. There is also a need for natural language programs that render ADP more useful to the physician.

- Because of the need to use the physician in areas where his education and training are most beneficial to the sick, there must be increased usage of para-medical and health care personnel for obtaining patient historical and laboratory data for input to the computer

memory. Since such technicians are practically non-existent, training and teaching programs must be developed. As the shortage of medical men is growing, there should be a gradual shift of emphasis from certain diagnostic functions toward the direction and management of therapeutic functions. Computers and para-medical personnel could perform more of the diagnostic functions, providing there are programs for the training of such para-medics.

- Many medical computer programs have been developed with Federal funds. These, then, should be in the public domain and should be readily available to hospitals and medical schools. In theory, they are available, but in actual fact, they cannot be found. Some method must be found to catalog and publicize the sources of these programs and to give them wider distribution.

- Almost all the medical schools and practically every hospital with 200 or more beds have data systems. In the areas of basic, as well as clinical, research, there is a good deal of duplication, which would be valuable if it were possible to compare results. The requirement, then, is that efforts should be made to coordinate the data activities among the medical schools and hospitals and perhaps create data networks so that the total data resource may be evaluated at one time in one place. The results would then be truly meaningful.

A major factor underlying these problems and recommendations suggested by the teaching institutions is that, regardless of the size of the hospital staff, very few of the clinicians engaged in research take advantage of the automatic data processing facility. There is a shortage of trained personnel to manage the program clinical studies

and a natural reluctance on the part of physicians to undertake and learn data processing and programming. Physicians spend many years of study and much effort to acquire professional efficiency; there is a shortage of medical personnel, and furthermore, the physician's capabilities are required where they will accomplish the greatest good for the largest number of people requiring medical attention. That is why there are usually less than 100 clinicians who use the electronic data processing facility as an aid to their research in any hospital no matter how large.

The survey analysis conducted in connection with this study brought forth two principal suggestions by the biomedical community which should be considered in determining directions for further effort:

- Detailed intensive studies should be conducted to determine what data physicians use in the decision-making processes associated with each of their activities; and

- A careful analysis should be made of the value and use of the data contained in the present standard medical record.

Pursuance of these tasks would lead to enhancement of the data flow from biomedical data generator to user.

SELECTED BIBLIOGRAPHY

1. Shannon, James A., Director, National Institutes of Health. Testimony before the Subcommittee on Government Research. March 1, 1967.

2. Chin, P.H., Computer Applications in Clinical Decision Processes. I B M Corporation, Mohansic, New York.

3. Magraw, R.M., and Magraw, B.B., Ferment in Medicine, W.B. Saunders & Co., Philadelphia, Pa., 1966, pp. 214-221.

4. Pfeiffer, J.E., "How Technology Advances the Art of Medicine," Think, July-August 1964, p. 6, I B M Corporation.

5. Engle, H. Martin, M.D., Chief Medical Director, Veterans Administration, Washington, D.C., "Research in the Service of Man: Biomedical Development," Excerpts from Testimony before U.S. Senate Subcommittee on Government Operations, March 2, 1967.

6. Mediros, A.A., "Workshop Notes," American College of Physicians Meeting, San Francisco, California, April 1967.

7. Collen, M.F., Rubin, L., Neyman, J., Dantzig, G.B., Baer, R.M., and Siegelaub, A.B.: "Automated Multiphasic Screening and Diagnosis." American Journal of Public Health, 54:741, 1964.
   Collen, M.F., Rubin, L., and Davis, L.: "Computers in Multiphasic Screening," Chap. 14 in Stacy, R.W., and Waxman, B.D.: Computers in Biomedical Research, Vol. 1, Academic Press.

8. Collen, M.F.: Periodic Health Examinations, Using An Automated Multitest Laboratory." JAMA, 195:830, 1966.

9. Collen, M.F.: "Automated Multiphasic Screening," a chapter in Sharpe, C.L.: Screening Techniques, Pitman Medical Publishing Co., Ltd., London (in press).

10. Collen, M.F., "Computer Medicine", Minnesota Medicine, 49:11 (Nov. 1966), pp. 1705-1707.

11. Collen, M.F., "Periodic Health Examinations Using an Automated Multitest Laboratory." JAMA: 195:10, March 10, 1966.

G. Pharmacology

1. Introduction

Pharmacology is that branch of the biological sciences which elucidates the reactions of foreign substances on animal cells, tissues, organs, and body systems. Pharmacological efforts are directed toward determining the activity of substances in humans. Veterinary pharmacology is directed toward similar studies in domestic animals. Toxicology and psychopharmacology are sub-disciplines which devote themselves to restricted segments of the total discipline. (Figure II-G-1)

Psychopharmacology is the study of the effects of chemical substances on normal and abnormal behavior. Modern research and development of drugs affecting behavior were greatly stimulated by the re-discovery of an ancient herb remedy, reserpine. A number of active substances were isolated from the herb, pharmacologically tested, and are, at this time, being successfully used in the treatment of various nervous and mental disorders. Subsequently, a rather large number of synthetic chemical compounds were developed. Some of these were found to be beneficial, and are now used in many institutions to assist ambulatory patients.

The knowledge, explanation, and description of the deleterious, noxious, or harmful effects of substances on living biological organisms is the purview of toxicology. This pharmacological discipline studies exposure, clinical, and biological effects, and describes the means for preventing and treating untoward reactions. Toxicity or poisoning may be due to drugs, cosmetics, foods, household chemicals, pesticides, air and water pollutants, agriculture, and industrial chemicals, or any other mixtures or compounds which may endanger animal or human health or life.

The systematic study of the poisonous properties of food and drink, and the study of effects, using animals as test subjects, began about 200 years ago. Later, it was recognized that all toxic effects are not apparent in acute experiments. Therefore, chronic, long term studies were introduced to determine the cumulative effects.

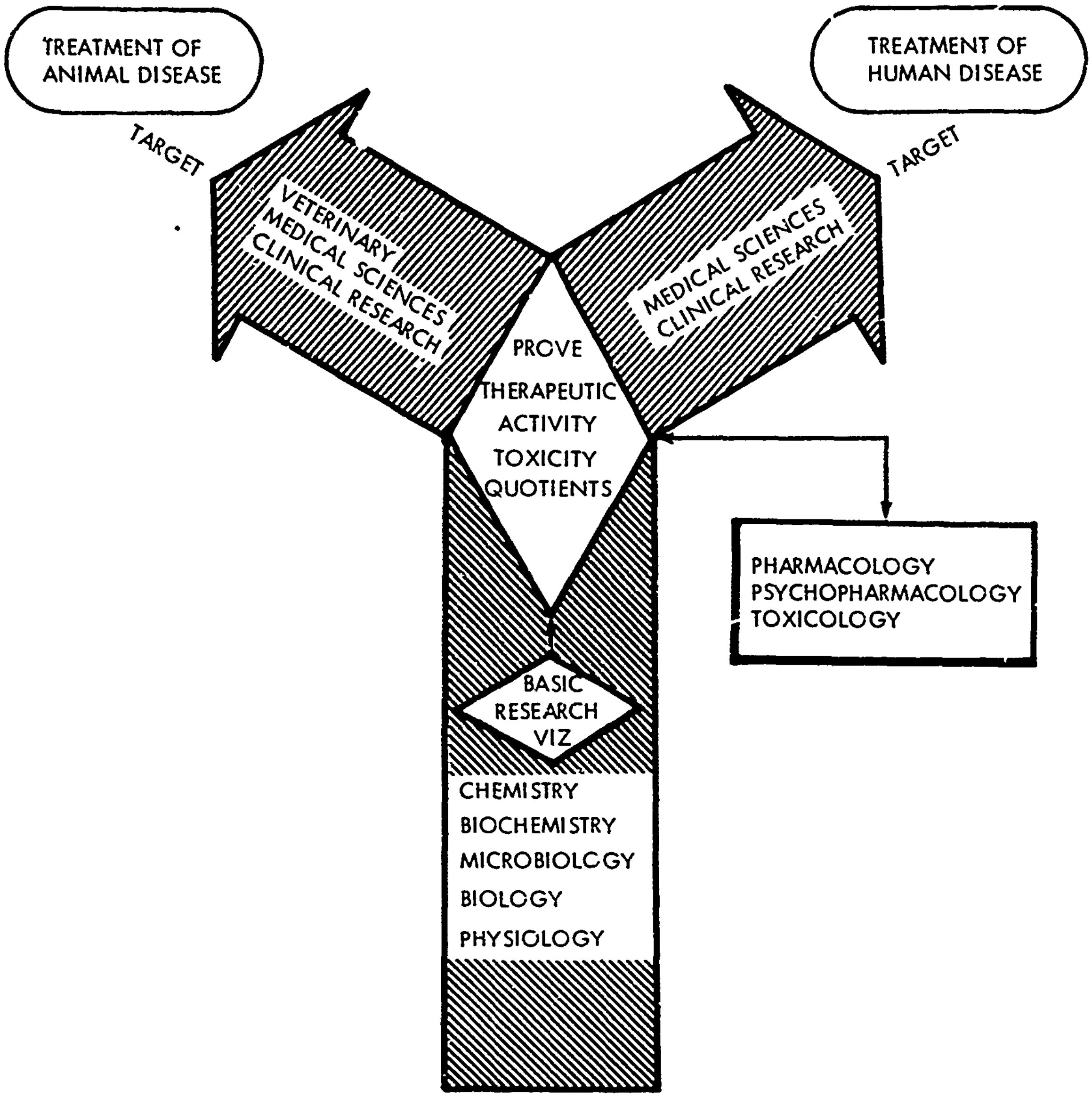


Figure II-G-1. Critical Position of Pharmacology Within the Biomedical Sciences
Note: All botanical, biological and chemical substances must pass pharmacological examination before they are permitted to be evaluated by the clinic.

Toxicological data are of the utmost importance to every segment of our population. Many very ordinary substances become dangerous when handled carelessly. Common household products have frequently caused injury and even death when incorrectly used or accidentally swallowed. Figure II-G-2 shows data classification within pharmacology.

2. Data Characteristics

There are a number of definitions for the term "data." These range from incisive, concise descriptions which state that data is knowledge expressed in digital or graphic form to those which connote much broader concepts. However, it seems that the usual characteristics are not applicable to pharmacological data. Although a certain portion is numerical in character, as when the drug dosage is expressed by weight, by far the largest amount must be expressed by means of short, factual descriptions. The following short sentence will serve as an example, "Atropine sulfate is an organic compound having mydriatic activity which can be easily observed when it dilates the pupil of the eye."

Basic Chemical Data. The pharmacologist has some interest in the chemical composition of the compounds he wishes to study. He desires to know the chemical structure of compounds, as his research may assist in uncovering, for the organic chemist, data on biological-activity-structure relationships. As further identification of the compound, he wishes to know the correct chemical name (I.U.P.A.C. or C.A.), the trivial, generic, or U.S.A.N., and also the trademark.

Basic Biological Data is of primary interest to the pharmacologist. He wishes to know everything about the macroscopic and microscopic effects on cells, tissues, organs, and the entire system. He examines normal cells and compares these with similar cells which have been infused with a solution of the compound he is studying. He often retains the slides, and he also may photograph the best ones. Strips of tissue are placed into a solution of the material under test and connected to an apparatus which records the tissue reactions. These graphs become a valuable portion of the experimental data. Changes in blood cells, nerve tissue, muscle, and blood vessels are observed. Data derived from these studies are stored as preserved slides, microphotographs, drawings, and written reports.

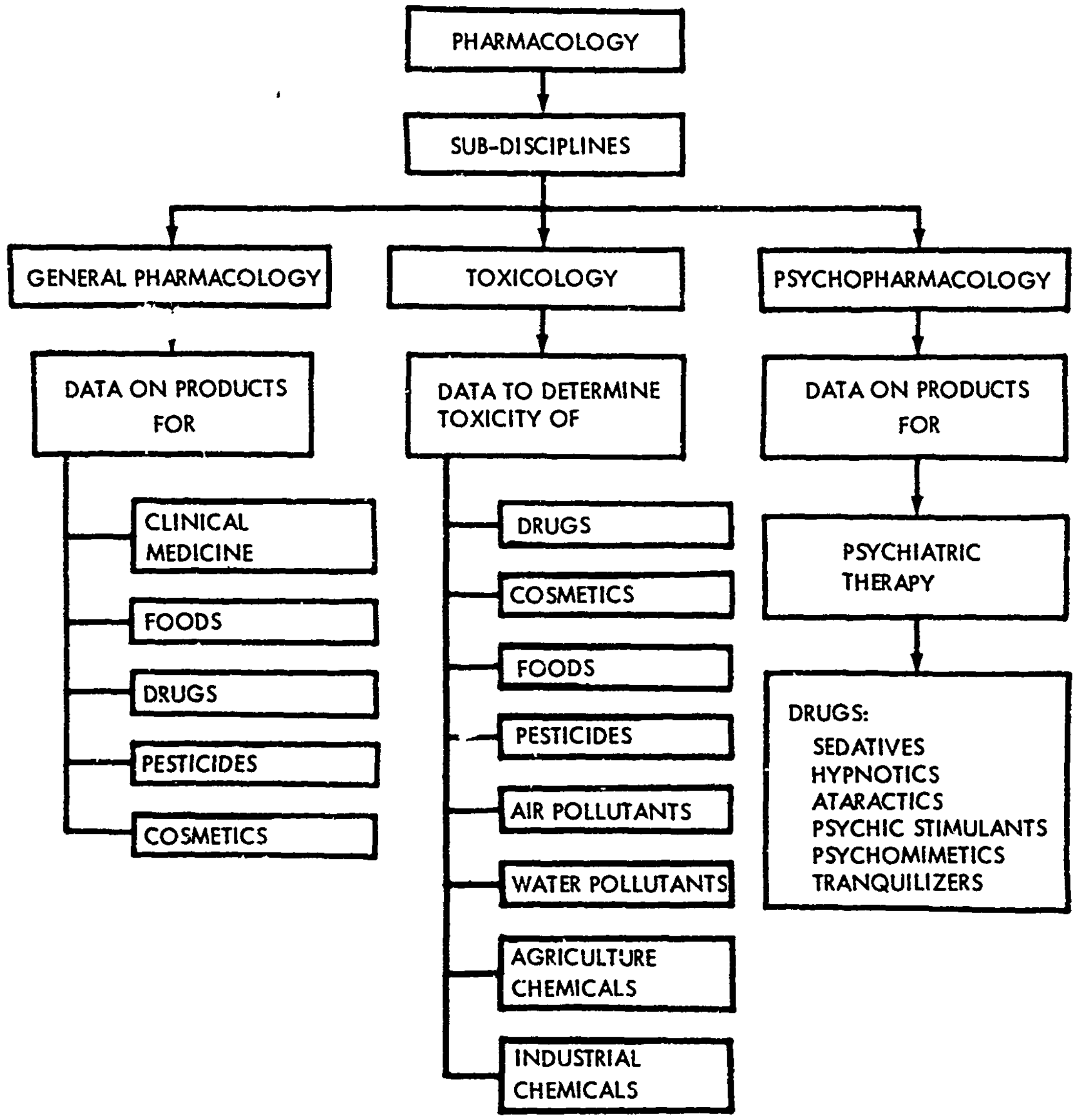


Figure II-G-2. Shows Data Classification Within Pharmacology. The Classification is Based Upon Sub-Field of Pharmacology. Commodities are Aligned Within the Specific Sub-Disciplines.

Toxicological Data are characterized by experiments which determine the hazards of exposure to drugs, cosmetics, industrial and agricultural or any other commodity which may cause an adverse reaction. Volatile compounds are tested by exposing the test animal to various concentrations. The smallest concentration which kills becomes the Minimum Lethal Dose (M. L. D.). The data are recorded as X parts per million. Similar experiments are conducted with liquids and solids. These may be administered orally, intravenously, intramuscularly, or interperitonally. The former and the latter tests are conducted on groups of animals belonging to the same species and specifically bred for biological test purposes. The tests for liquids and solids are recorded as L.D. 50's, meaning the smallest lethal dose which kills 50% of the test animals. Some investigators record L.D. 100's or the smallest dose which kills all the test animals. Important to all toxicological data are the acute studies wherein the determinations are made for the smallest single lethal dose and the chronic studies wherein determinations are made to find the biological effects of various dosages. Chronic, long term studies determine organ and tissue damage, if any, as well as mortality. Recorded data contain the numbers of animals used in the experiments, various dosages administered, time factors, diets, reactions to the compounds, and statistical analyses.

Psychopharmacological Data annote the effects of chemical substances upon normal and abnormal behavior. Behavioral patterns are carefully studied and recorded. Studies are conducted to determine the modifications in the structure of the brain and nervous system which may be the cause of behavioral changes. These are both macroscopic and microscopic. The recorded data may be preserved as photographs, slides, or carefully documented reports. Anti-psychotic drugs, when used clinically, are of tremendous economic importance, for they have enabled many ambulatory, emotionally disturbed persons to resume a relatively useful life. Treatment of many institutionalized patients has been sufficiently successful to permit their release from psychiatric hospitals.

Pharmacological data are used as the basis for many research and development decisions which literally may mean the difference between life and death. For the decisions to conduct clinical studies on people are made as a result of careful studies of these data.

Drug Data. The aim of all pharmaceutical research is to develop a product which advances the treatment of a disease. During the later stages of clinical research, the data on treatment are refined and organized for the benefit of the medical practitioner. The information relative to the practical therapeutic application of the substance becomes the therapeutic or drug data.

Drug data describe the therapeutic indications or the symptoms and type of disease wherein the drug is effective; the contraindications, or the clinical conditions where the drug must not be used, also where it may be used cautiously; the proper range of dosage to be administered; and the most suitable method and form of administration.

3. Data Flow

Data Sources. Pharmacological and toxicological information and data are generated in every part of the world. They are published in many languages and find their way to the user through many channels. (Figures II-G-3 and II-G-4)

Pharmacological data, including that of toxicology and psychopharmacology, are largely generated in the universities and hospitals. Research laboratories of the Federal and State governments also generate considerable amounts of data. Information and data from these sources are usually published in the open literature. By far, however, the largest amount of information and data on pharmacology and toxic effects can be found in the files of the chemical, pharmaceutical, and cosmetic manufacturers. These firms maintain very efficient, well staffed pharmacology departments. Naturally, certain companies cannot afford to support an activity of this magnitude, so they contract with private research organizations to conduct their biological-activity-toxicity studies.

Data Users. Every substance which touches, is ingested, or inhaled by a human being or a domestic animal causes physiological reactions. These may be beneficial or harmful. Since pharmacology studies the reactions, it becomes obvious that manufacturers of foods, drugs, cosmetics, chemicals, pesticides, household products, as well as those responsible for the welfare of groups of people, must and do have an enormous interest in pharmacological data. (Table II-G-1)

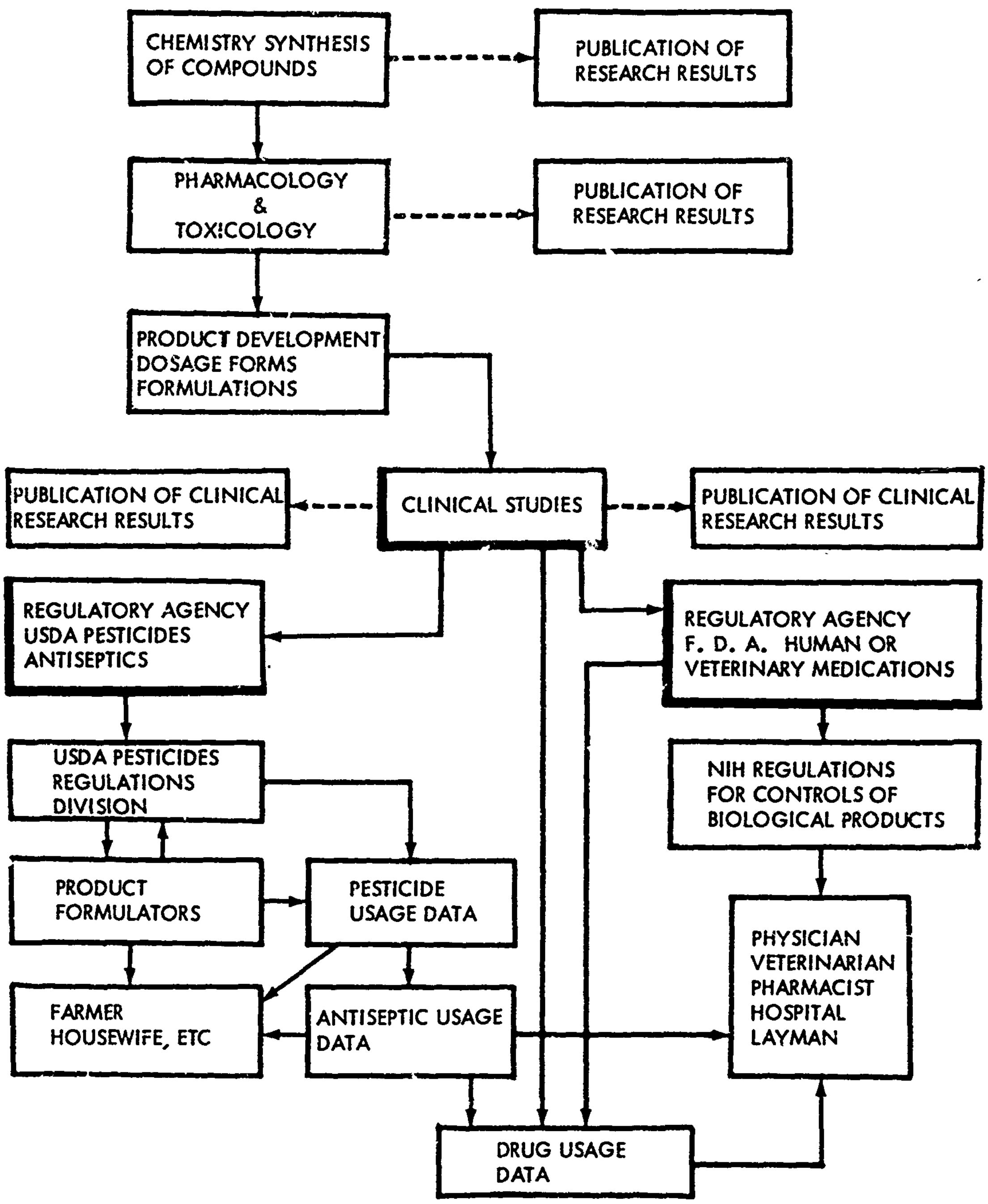
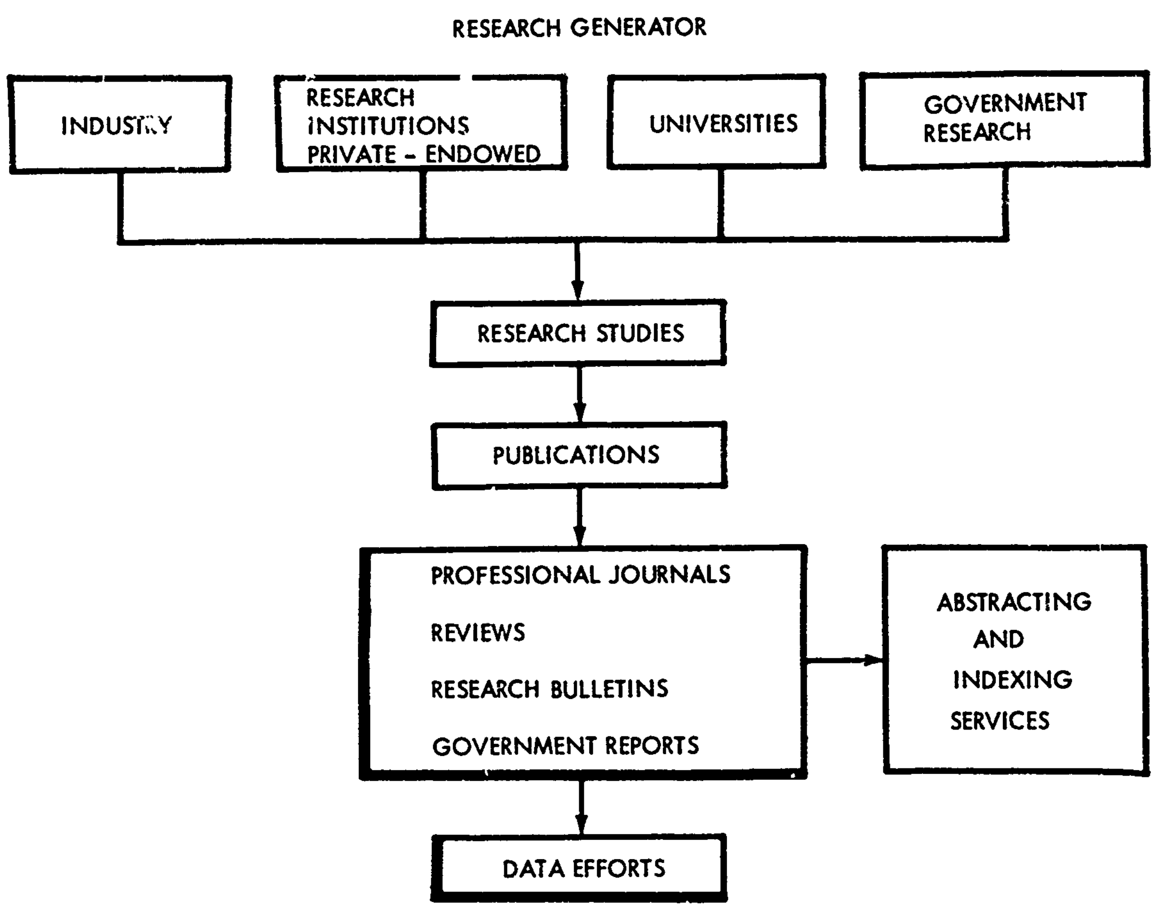


Figure II-G-3. Steps Toward the Development of a Biologically Active Substance. These Steps Determine the Development and Flow of Data Pertinent to Pharmacology.



NOTE: INSTITUTIONS AND ACTIVITIES INVOLVED IN THE GENERATION AND PUBLICATION OF PHARMACOLOGICAL DATA.

Figure II-G-4. Generation and Publication of Data in Pharmacology

TABLE II-G-1. TYPICAL USERS OF DATA GENERATED BY EACH OF THE MAJOR SUB-FIELDS OF PHARMACOLOGY

General Pharmacology

Physicians
Veterinarians
Clinical investigators
Food manufacturers
Pesticide producers
Pharmaceutical manufacturers
Cosmetic manufacturers
Chemical manufacturers
Veterinary medical manufacturers

Psychopharmacology

Psychiatrists
Psychiatric nurses
Police departments (psychologists, psychiatrists, medical officers)
Drug addiction officials
Public school systems (psychiatrists, psychologists)

Toxicology

Physicians
Veterinarians
Nurses
Medical examiners (coroner's office)
General practitioners
Hospital officials
Prison officials
Police officials
Social workers
Government organizations
Armed services

Obviously, then, manufacturers of all sorts of commodities must consult the pharmacological and toxicological data before undertaking any distribution and marketing programs. Commercial firms therefore retain, as staff or as consultants, qualified scientists who evaluate the data for the management. Foods, drugs, pesticides, biologicals, antibiotics, and antiseptics must be submitted to government regulatory agencies for approval before the products may be marketed; pharmacological data must accompany all applications. This assists government scientists in arriving at the correct regulatory decisions. (Figure II-G-5)

Many government and non-government agencies require toxicological data for the prevention of deleterious effects upon their personnel and the public. Figure II-G-6 graphically portrays these groups.

Due to the proprietary nature of the data contained in the files of chemical, pharmaceutical, food, and cosmetic manufacturers, this large quantity of information has not been, nor will it ever be, published. While it is available only to certain employees of each company under certain, discrete circumstances, a great many firms will share their data with qualified persons who do not have a conflict of interest. Each inquiry receives individual study, and when a favorable decision is reached, the scientist is permitted access to the data. This finding especially applies to the pharmaceutical and chemical companies.

The medical schools and the funded research institutions permit ready access to their data files. Of course, the inquirer must be a qualified scientist. Certain Federal research programs of a classified nature generate pharmacological-toxicological data which are unavailable for other uses.

The position of pharmacology within the biomedical sciences is most critical. Before any chemical, biological, or botanical substance is given to the clinician for trial on patients, it must pass critical pharmacological tests. These determine the compound's activity in the various organs, the toxicity, if any, and, in the case of drugs, the probable dosage levels. (Figure II-G-1)

Certain government regulatory agencies, notably the F.D.A. and the U.S.D.A. Pesticide Regulation Division, require extensive pharmacological-toxicological data. The F.D.A. needs the data to determine if the products are effective and safe for human or animal use. The

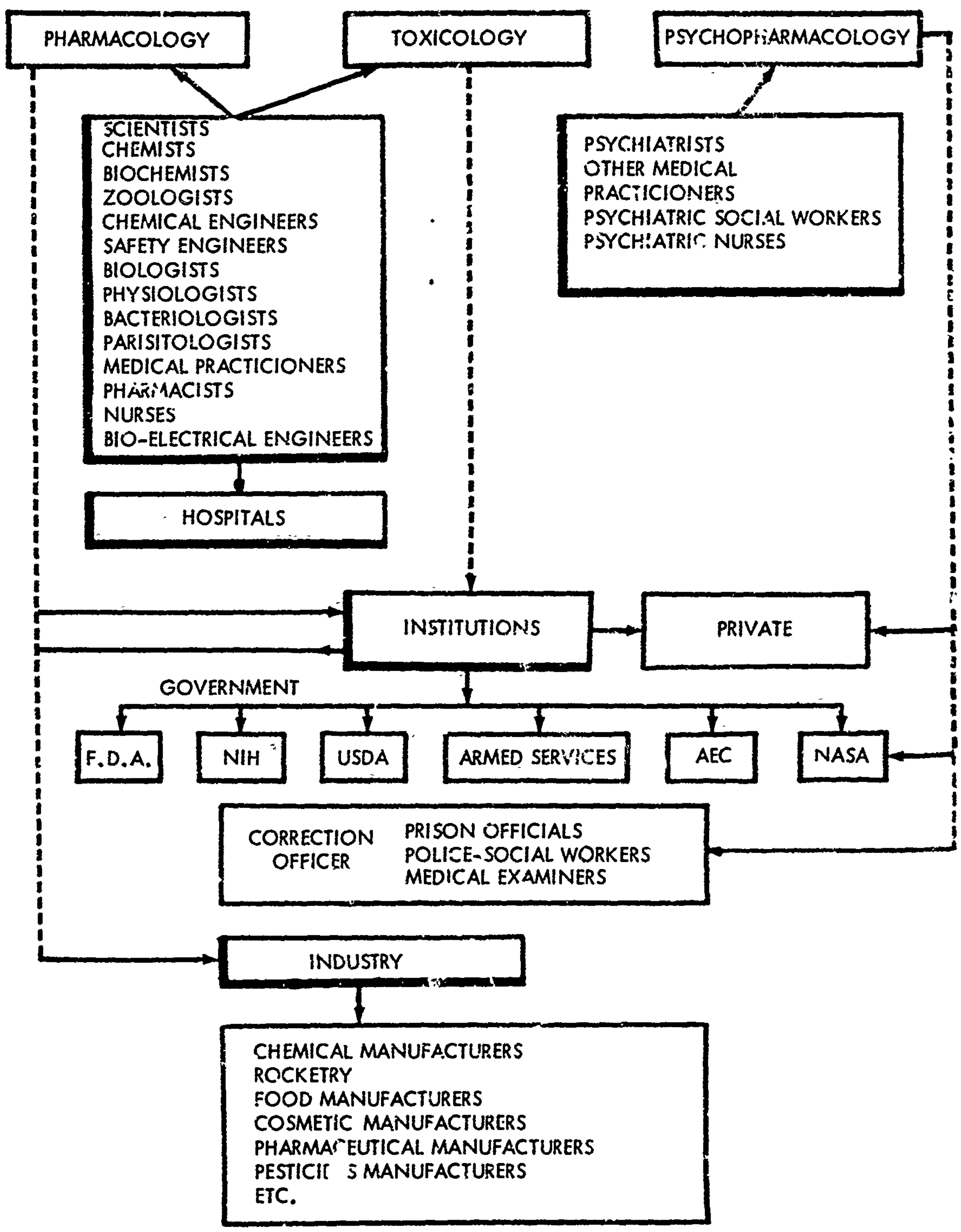
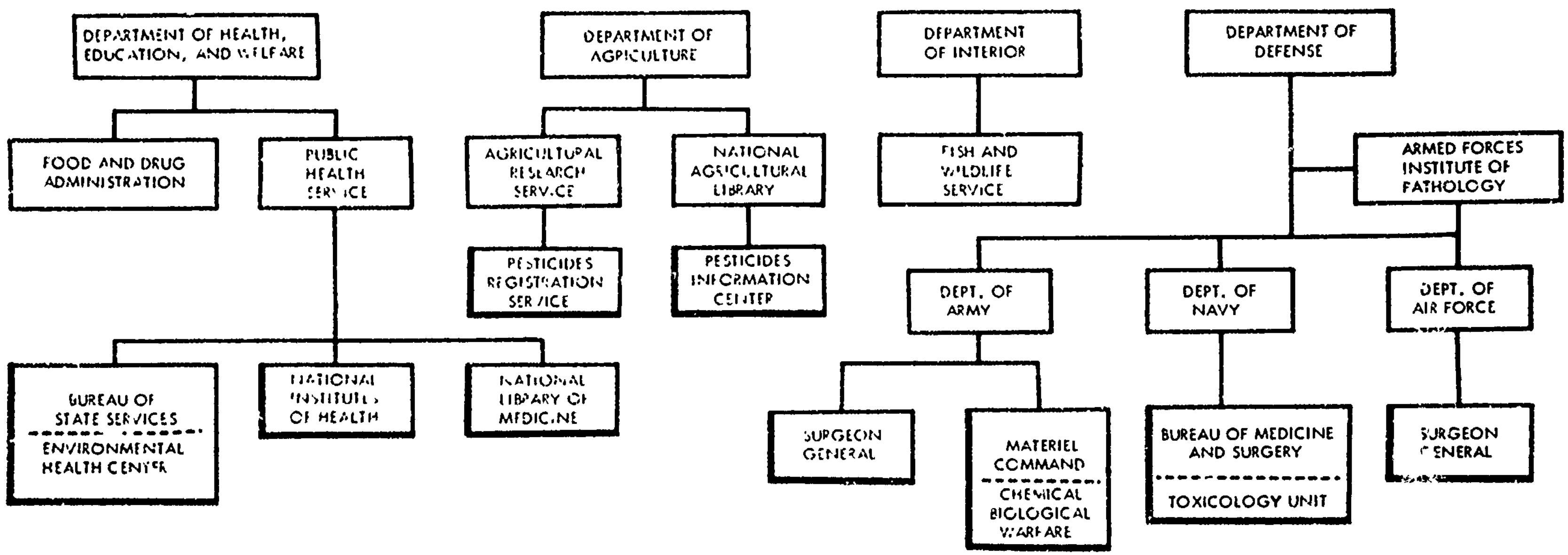


Figure II-G-5. Users of Pharmacological Data by Institutions and Occupational Groups



* NON-GOVERNMENTAL ORGANIZATIONS WITH SUBSTANTIAL ROLES IN TOXICOLOGICAL INFORMATION

1. NATIONAL ACADEMY OF SCIENCES NATIONAL RESEARCH COUNCIL-ADVISORY CENTER FOR TOXICOLOGY, DRUG RESEARCH BOARD.
2. AMERICAN MEDICAL ASSOCIATION-DIVISION OF DRUG SAFETY, COUNCIL ON DRUGS
3. PHARMACEUTICAL MANUFACTURERS ASSOCIATION-COMMISSION ON DRUG SAFETY EVALUATION
4. AMERICAN PHARMACEUTICAL ASSOCIATION AMERICAN SOCIETY OF HOSPITAL PHARMACISTS
5. AMERICAN SOCIETY FOR PHARMACOLOGY AND EXPERIMENTAL THERAPEUTICS
6. SOCIETY OF TOXICOLOGY

*DOES NOT INCLUDE THE NUMEROUS PUBLISHING, ABSTRACTING, INDEXING ORGANIZATIONS. ABSTRACTED FROM: HANDLING OF TOXICOLOGICAL INFORMATION, A REPORT OF THE PRESIDENT'S SCIENCE ADVISORY COMMITTEE, WASHINGTON, D. C. JUNE 1966

Figure II-G-6. Federal Government Organizations with Substantial Roles in Toxicological Information*

Pesticide Control Division requires the data for the correct label directions on pesticide packages and to insure that the user shall receive instructions which will prevent accidental poisoning. All such data belong to the manufacturers. These agencies recognize the proprietary rights of the owners and therefore, the information is not for publication or use by unauthorized personnel. (Figure II-G-7)

The veracity of a great deal of pharmacological-toxicological data has been authenticated over many years. For the newer products, however, the clinical pharmacologist is constantly on the lookout for the unexpected and unknown side-effects. Sometimes these are not discovered until the preparation has been used, exposed, or administered to large numbers of people for many years. Obviously, when such incidents occur, the data are immediately corrected. However, it may be postulated that the data on products used for many years have little chance of becoming obsolete, but the data on new preparations may be corrected at any time. It may be remembered that, unlike the engineering field, where hard data remain virtually unchanged, human reactions are subject to changes resulting from a variety of causes, such as different diets, changes of climate, metabolic disturbances, or emotional upsets. These, as well as other factors, will modify the responses toward compounds, thus necessitating modifications or changes in the related data.

Pharmacological and toxicological data systems offer a number of primary index points of entry. The data can be locatable by searching for a specific chemical, a group of related chemicals, an animal species upon whom a series of tests were performed, a biological reaction, an organ system, a metabolic change, a common, trivial, generic, or trade name. Data are stored and retrieved in several of the usual forms. Storage is on simple file cards, or one of the several punch card systems, or data may be programmed and fed into a computer. Several activities are now engaged in planning and coordinating pharmacology data; the following are examples:

- The Drug Information Association was organized to further the modern technology of communication in the medical, pharmaceutical, and allied fields. It provides a climate of cooperation in order to expedite the transfer of drug

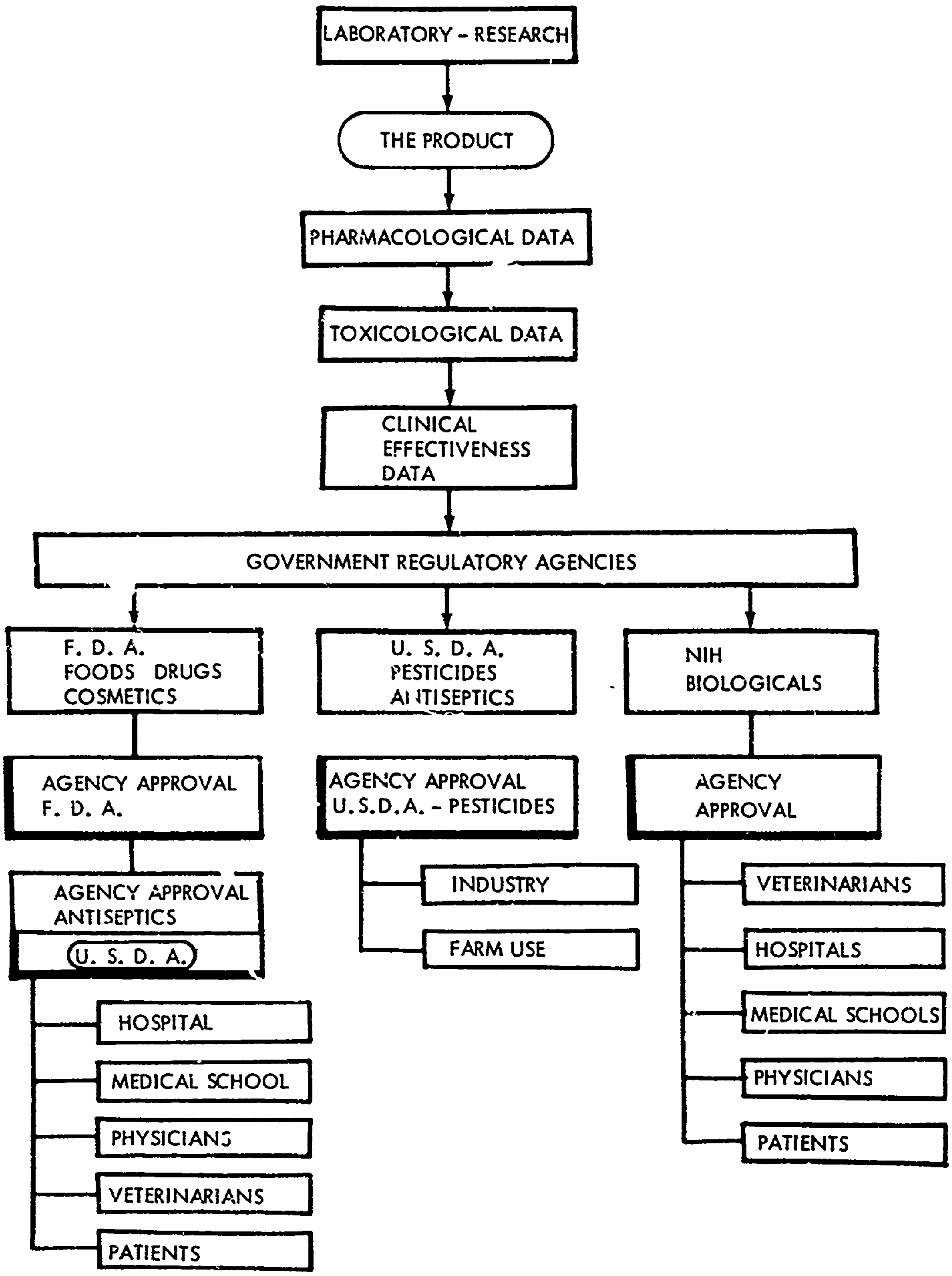


Figure -G-7. Illustrates Where Government Regulatory Agencies Interdict the Flow of Pharmacological Data from the Developing Laboratory to the Consumer

information from the data generators to the data users with a minimum of duplication of effort. The association publishes a quarterly bulletin, holds meetings and seminars, and encourages exchange among its members.

- The American Medical Association, Department of Drugs, maintains an active registry and files on drug data. There are over 30,000 items in this bank. Any qualified person may request data on a specific product. It will be answered by letter and, if the answer contains data and information of value to the entire medical profession, it will be published in the Journal of the American Medical Association. All the data and monographs are annually collected and published in book form as New Drugs.

- National Academy of Sciences is conducting studies for the F.D.A. of drug efficacy in humans and domestic animals. Approximately 1,100 veterinary drugs will be studied under the aegis of a 12-member committee.

- Drug Research Board was formed in 1963. The membership is composed of internists, pediatricians, pharmacologists, and toxicologists. Members of the Board come from industry, government, and the universities. The Board attempts to survey the policies, principles, and practices which influence drug research, and to provide opportunities for discussion of the problems of investigative medicine, industry, and government. The Board cooperates with the American Medical Association, the Pharmaceutical Manufacturers' Association, and the Food and Drug Administration.

In addition to these and other planning and coordination activities, there are several formal data efforts, such as networks, data publishing programs, data centers, and data-document depositories. The following pages describe the primary formal data efforts in the pharmacology field.

Among the data networks, two examples best illustrate the nature of the technical and data handling activities involved:

- The Committee on Drug Addiction and Narcotics of World Health Organization (WHO). The National Academy of Sciences manages, under the sponsorship of WHO and the Veterans Administration, coded data and information on narcotics and addiction. The collecting, indexing, and coding of pertinent literature are carried out by the committee in collaboration with the American Social Health Association, Inc. and the Alcoholic and Drug Research Foundation (Toronto). These groups prepare the index of the literature on addiction, addicting drugs, analgesics, and antitussives. It is coded on master cards, according to: drugs; categories (characterization of materials); effects; addiction; habituation and tolerance; and modifying factors. The coded master cards are sent to Smith,Kline and French, which transfers the coded information, essential bibliographic information, and the accession numbers of the master cards to I B M cards, in preparation for computer searches of the index. The master cards are microfilmed in the order of their accession numbers. Copies of the microfilm and the I B M cards are deposited with the committee, the American Social Health

Association, Inc., the Alcoholic and Drug Addiction Research Foundation, the Department of Pharmacology of the University of Michigan, and WHO in Geneva. The committee was responsible for publication, in 1941, of a complete review of the pharmacological literature in the field of opium alkaloids. The publication contains about 10,000 items arranged chronologically.

- The F.D.A., Bureau of Medicine, Drug Information System is another data network. This is an alerting system designed to detect previously unknown untoward effects of drugs or the incidence of adverse effects which were greater or less than the previous experience with a limited number of patients. An adverse reaction is defined as one which is noxious, unintended, and occurs at doses normally used in man for the prophylaxis, diagnosis, or therapy of disease, or for the modification of a physiological function.

  Input data are taken from reports sent in by hospitals, private physicians, the drug industry, A.M.A., and the professional staff of the F.D.A., Bureau of Medicine. The items are keypunched on cards and entered on magnetic tapes. The data consist of all the diagnostic factors, pharmacology, toxicology, reaction factors, drug sources, and the outcome of the cases. These data measure the adverse effects of drugs. The file grows at the rate of several thousand items per month.

Pharmacological Data Publishing. A recent study (1965-1966) of the world's serial literature in pharmacology, toxicology, and cosmetology identified 1,066 current serials, 814 of which devoted at least 50 percent of their pages to original research in these fields. There are additional thousands of journals with a lesser fraction of this type of contribution.

Journals containing pharmacological-toxicological data are included in many diverse fields, such as agriculture, biology, chemistry, engineering, nuclear science, and psychology. The major journal language is English. Of the total number of major journals, 1,066, English was the language in 536. Other languages used in the major group of serials were: French--173, German--160, Spanish--112, Japanese--86, Italian--53, Russian--42, Portuguese--37, and 23 in all other languages.

It has been estimated that there are 200,000 to 300,000 original papers published each year which contain drug-oriented literature of value and interest to the biomedical scientist. The magnitude of the data problem is indicated by the fact that approximately 75,000 new chemical compounds are reported in the literature each year. However, the pharmacological and toxicological data are reported only for those compounds which are of potential value as therapeutic agents or where exposure to industrial compounds may be hazardous to man. While the serial publications are most important, there are a number of very useful bound books and services of great value to pharmacologists, toxicologists, medical practitioners, hospital pharmacists, and retail pharmacists. Representative publications in the different areas of pharmacology are listed as follows:

> Excerpt Medica, Section II, Physiology, Biochemistry, and Pharmacology: Herengracht 119-123 Amsterdam, Netherlands, and 2 East 103rd Street, New York, New York 10029. A monthly service excerpting the most important articles in the world's literature.
>
> The Pharmacological Basis of Therapeutics: Goodman, Louis A., and Gilman, Alfred; W.B. Saunders, 1965 (Handbook).

Pharmacology in Medicine: Drill, Victor A., McGraw-Hill and Company, New York, New York, 1965 (Handbook).

Pharmacological Reviews: American Society of Pharmacology and Experimental Therapeutics, International Reviews of pharmacology and experimental therapeutics, proceedings of international meetings.

The Annual Review of Pharmacology, Annual Reviews, Inc., Palo Alto, California.

The Pharmacologist, semi-annual publication of American Society for Pharmacology and Experimental Therapeutics.

Clinical Toxicology of Commercial Products: Gleason, M.N., Gosselin, R.E., and Hodge, H.C., Williams and Wilkins, Baltimore, Md. This is a practitioner text, with supplements distributed to subscribers in health departments, hospitals, industrial companies, medical schools, etc.

Handbook of Toxicology, Vol. I: Acute toxicities of solids, liquids, and gases, Spector, Wm.S., ed., W.B. Saunders Company, 1956.

Handbook of Toxicology, Vol. II: Antibiotics, Spector, Wm.S., ed., W.B. Saunders Company, 1957.

Handbook of Toxicology, Vol. III: Insecticides, Negherbon, Wm.O., ed., W.B. Saunders Company, 1959.

Handbook of Toxicology, Vol. IV: Tranquilizers, ed., Grebe, R.M., W.B. Saunders Company, 1959.

Handbook of Toxicology, Vol. V: Fungicides, Dittmer, D.S., ed., W.B. Saunders Company, 1959.

Drugs of Choice, Modell, W., C.V. Mosby and Company, St. Louis, Missouri, 1960.
Gives all the data necessary for treatment.

Modern Drug Encyclopedia and Therapeutic Index, plus monthly supplements "Modern Drugs." Drug Publications Division, Reuben Donnelly Corporation, New York.
Source for information on U.S. proprietary drugs. Material is arranged by trade names, giving action and uses, administration, supply, and contraindications. Indexes of manufacturers and distributors with products, and by general subject including generic names.

Facts and Comparisons, Kastup, E., and Schwaca, G., Facts and Comparisons, Inc., St. Louis, Missouri. Arranged by groups of products and use. Permits easy comparison of similar or related products with common ingredients, actions, side effects, and contra-indications. A cost index permits the comparison of the cost of two or more comparable products. Additional new product listings issued each month.

Physicians' Desk Reference, Medical Economics, Inc., Ordell, New Jersey.
An annual with quarterly supplements. Assists physicians to keep pace with progress and introduction of new pharmaceutical specialties, biologicals, and antibiotics. Sections are arranged by brand name, company, generic name, therapeutic indications, and major ingredient.
A professional products section gives detailed new product information on composition, action dosages, and forms.

In addition to these publishing activities, the following services are provided which disseminate data contained in the published literature:

deHaen Drug Index Data Systems. There are four deHaen Drug Index data card systems. Each serves a slightly different purpose. They are interlinked through therapeutic and pharmacologic classifications, code numbers, and non-proprietary names. One card will automatically lead to a card on the same product in another system. This type of indexing provides a comprehensive record of literature analysis of a specific drug and similarly used components.

The deHaen data center is located in New York City. It is a private enterprise, operated for profit by an in-house staff.

All the data are abstracted from published material and other research journals. They are formatted on 5" X 7" cards, cross indexed for retrieval, and suitable for transfer to a machine-automated system.

Drugs in Prospect. This is an alerting service designed for chemists, pharmacologists, and literature scientists. The data evolve around new chemical compounds which may develop into useful drugs. The cards contain the therapeutic classification of the drugs, pharmacological categories, place of origin, chemical composition, results of pharmacological tests, chemical structure, accepted (CA) nomenclature, and references. Additional information is added as it appears in the literature. The service supplies about 2,000 cards per year.

Drugs in Research. This data service is directed primarily to the librarian, the literature scientist, and the market research man, and only secondarily to the practicing scientist. It supplies a monthly cumulative bibliography and other data covering products currently known to be in the research

stage in the United States and other countries. Marketing information on the status of the product in other countries, authors, names, and titles of the papers are also portions of the data. The service supplies about 1,600 data cards per year.

Drugs in Use. These data cards are analyses of the clinical literature. Literature abstracts are transformed into a unique type of format wherein the data are easily recognized and retained. These data can be transferred into automated systems for analysis of details by: product, class of drugs, age and sex of patients, dosage, concomitant therapy, diagnosis, clinical results, and adverse reactions. There is a standard vocabulary available for the last three items (diagnoses, clinical results, adverse reactions) so these data can be computerized for automated analysis. Drugs in Use covers approximately 5,000 cards per year. The data are abstracted from clinical papers, published in 410 journals originating in 22 countries.

The subscribers to the deHaen services fall into these classes:

Chemists
Pharmacologists
Clinical Investigators
Practicing Physicians
Hospital Pharmacists
Government Agencies (NIH, HEW, FDA)
Medical Librarians
Science Information Departments
Market Research Departments
Medical Writers

Data Centers. Practically all the pharmacological data centers are specialty-oriented activities. This seems to be true, whether they are operated by the government, industry, a research or teaching institution. Pharmacological centers contain data on either drugs,

cosmetics, foods, pesticides, or clinical medicine. Psychopharmacological centers contain data on drugs used in the treatment of psychiatric disorders. Their interests lie with drugs such as sedatives, hypnotics, ataractics, psychic stimulants, and psychomimetics. The toxicological centers are interested in the poisoning propensities of drugs, cosmetics, foods, household chemicals, pesticides, air and water pollutants, agriculture and industrial chemicals, or any other mixtures or compounds which may endanger animal or human health and life.

The Federal Government operates several important national data centers. These are the NIH National Clearing House for Mental Health Information (Psychopharmacology), NIH Cancer Chemotherapy National Research Center, NIH Heart Institute, and the U.S. PHS Poison Control Branch. The latter distributes data and information to the poison control centers throughout the country. The following are descriptions of typical data centers:

- Cancer Chemotherapy National Service Center - NIH. This is a data processing facility having more than 180,000 chemical compounds in the registry. New compounds are screened at the rate of 50,000 per year and the pharmacological data are recorded and reported. The automated data processing system handles approximately 1,000 test reports daily from drug screening laboratories on the effects of drugs on animal cancer tumors. Practically all data are available to qualified research workers. The exceptions are data on compounds whose proprietors object to disseminating the data.

- Psychopharmacology Service Center - NIH. This center was established in 1956 by the Institute of Mental Health to stimulate research on drugs which might improve mental health. The center indexes, codes, compiles, and analyzes published and unpublished data and information on psychopharmacology. Its collections contain approximately

18,000 books, reprints, reports, and abstract journals, with identifying information indexed and coded for machine retrieval. Scientists and clinical psychopharmacologists may obtain copies of Psychopharmacological Abstracts, without charge. It contains the latest data and information. The staff will provide data on specific drugs and data, as well as technical consultation ahd help in the development of drug studies.

- Medical Records - NIH. The Medical Records Department of the NIH Clinical Center compiles and processes clinical data for use in administrative research and in studies of diseases under investigation. About 30,000 entries are transcribed each year relating to the conditions of patients participating in clinical research studies. These include medical histories, results of examinations, records of treatments and operations. Statistics are analyzed for use in current reference or for retrospective searching.

- National Clearinghouse for Mental Health - NIH. This is a central repository for data and information on all aspects of mental health. The clearinghouse exchanges data and information with other Government and State agencies, national organizations, voluntary groups, professional societies and universities. It develops and supplies specialized data and information for scientists, lay groups such as correctional officers and police departments, and the Institute's staff. It serves as a national referral center for inquiries in the field of mental health.

- Section on Medicinal Chemistry - NIH, Laboratory of Chemistry -- NIAID. Serves as a clearinghouse for all new analgesic drugs received from pharmaceutical firms and individual investigators because of potential addiction liability of potent pain-relieving drugs. Members of this section maintain contact with the staffs of drug companies and with investigators who have submitted or proposed new analgesic compounds. The data obtained from preliminary investigations at the laboratory are transmitted to interested persons.

- Toxicology Data Centers. Toxicological data centers compile and disseminate descriptions of the poisonous or dangerous effects of chemicals, drugs, and cosmetics due to improper handling, exposure, overdosage, inhalation, skin absorption, and ingestion. An important portion of the data concerns methods needed for the prevention or relief of untoward, toxic effects.

- Poison Control Centers. The poison control centers which serve the general public deserve some special mention. These are autonomous organizations developed by local medical or paramedical groups in cooperation with the State Health Departments. Most of them are located in hospitals; the balance are in the health departments. They are organized to maintain information on the formulation and toxicity of the many products on the market and the treatment necessary to counteract any dangers due to accidental ingestion; to improve the treatment facilities of the hospitals so treatment may be expedited; and to establish a reporting system to obtain information on the cause of accidental ingestions so that preventive programs might be developed. In an emergency,

the local physician can obtain data and information on toxicity and treatment for a substance by contacting the local center. During 1966, this activity brought the total number of poison control centers authorized by state health departments to 550. The National Clearinghouse for Poison Control Centers, a unit of the Accident Control Division of the U.S. Public Health Service, provides the information necessary to determine the nature of the products being ingested and develops substantive approaches to prevent accidental ingestion.

The importance of the service may be understood easily. During 1964, as in preceding years, there were approximately 500 deaths from the estimated one-half to one million accidental ingestions of medicines and commercial household products.

Data-Document Depositories. Libraries are, naturally, an integral portion of every medical, pharmacy, dental, and veterinary college. Since pharmacology is an important subject, each institution has in its library a great deal of material for use by the students, professors, and researchers. The collections usually contain texts in English, French, German, and other languages, reprints of important papers, serials, and laboratory handbooks. Health, agriculture, and the armed services maintain libraries containing data, documents, reprints, journals, and texts on pharmacology. Certain collections, such as the one in the National Library of Medicine, are very extensive.

Industry has a great need for data and documentation in pharmacology. In addition to satisfying the government regulatory agencies' requirements for bibliographic data, industry requires information to assist in labeling, marketing, and legal matters, essentially to help document their position in malpractice suits and patent conflicts.

Industrial libraries are oriented toward the primary company needs. A chemical company wants to have all the information it is possible to obtain on toxicity. They also need all the data concerning different types of exposure to the products of their manufacture. Their library, therefore, will subscribe to journals likely to contain papers relating to their interests. Libraries belonging to food, drug, cosmetic, or veterinary product manufacturers are similarly motivated. The document collection in industrial libraries will be subject and product directed, as will be the reprints, journals, and reports.

Firms performing research and development for industry for financial rewards must maintain extensive libraries in order to satisfy client needs. The successful contract R&D laboratory usually has an extensive library containing some documents and references on all subjects, or it has ready access to a good one in a university or a specialized one, such as the Chemists' Club Library in New York or the John Crerar in Chicago.

5. Principal Problems in Pharmacological Data Management

Review of pertinent literature and exploratory interviews have identified the following candidate issues for exploration in workshops and questionnaires:

- Users of pharmacological data place too much credence in data from analytical chemical techniques and from screening tests in animals.

- There are deficiencies in the methods used in the publication of pharmacological-toxicological data which cause duplication of research.

- There is no center in the United States which confirms experimental data and establishes animal species norms. There is an apparent need for a pharmacological center, probably government-operated, which will cross-check the data and become the source of confirmed experimental data.

- References and handbooks in pharmacology are usually published with long intervals between newly issued volumes. With the exception of Clinical Toxicology of Commercial Products by Hodge, H. C., et al., there are no publications which are augmented by monthly or quarterly supplements.

SELECTED BIBLIOGRAPHY

1. duBois, K.P., and Geiling, E.M.K., Textbook of Toxicology, Oxford University Press, New York, 1959.

2. Koppanyi, T., and Avery, M.A., "Species Differences and Clinical Trial of New Drugs: A Review," Clinical Pharmacology and Therapeutics, 7:2 pp. 250-270, March-April 1966.

3. Handling of Toxicological Information, A Report of The President's Science Advisory Committee, Office of Science and Technology, June 1966.

4. Cohn, Gunther, and Carr, Carol, An Evaluation of Data Collections for Plastics, The Franklin Institute, May 1964.

5. "Use of Poison Control Centers," Address by: Henry Verhulst, Chief, Poison Control Branch, Division of Accident Prevention, U.S. Public Health Service, 29 October 1963.

6. Verhulst, H., Crotty, J.J., and Maisel, G., "The Pharmacist in Poison Prevention," Journal of the American Pharmacology Association; NS 6:1, January 1966.

7. Krueger, H.; Eddy, N.B.; and Sumwalt, M., Pharmacology of the Opium Alkaloids (Supplement to No. 165, Public Health Reports, Federal Security Agency, U.S. Public Health Service, Government Printing Office, 1941).

H. Social and Behavioral Sciences

1. Introduction

For the purpose of this study, the social and behavioral sciences are defined to include those involved in the scientific and technical aspects of administration and management; anthropology, documentation and information technology, human factors engineering, linguistics, personnel selection, training and evaluation, psychology and sociology. The principal emphasis of this subsection is the extent to which social and behavioral scientists have become aware of the inadequacy of the data bases within their fields; and, as a result, have begun to develop a variety of data banks of the large amount of available but heretofore inaccessible data that have been collected over the years.

In the social and behavioral sciences, it has been found that greater amounts and varieties of data are increasing in their usefulness to research. Accuracy of data has also become a more stringent requirement in research studies. Corresponding to the increased emphasis on data, researchers are discovering the utility of the computer as a new and important tool with which to manage the data. Dr. Harold D. Lasswell has stated that "The computer revolution has suddenly removed age-old limitations on the processing of information including the linkage of data with competing theories of explanation." To a large degree, the computer has created among the social and behavioral scientists an atmosphere in which the role of data is of increasing importance. It is now felt that the scarcity of computerized or machine-readable data impinges on the opportunity to extend immeasurably the body of knowledge in the social and behavioral sciences. To meet this challenge, many financial and manpower resources are now being invested to develop the necessary data bases in which a vast amount of data can be made more accessible to the social and behavioral sciences. The federal government, major universities, and state and local governments are establishing data banks and developing programs to make the data within their fields of interest accessible to administrators, managers, researchers, and those in other disciplines who have a need for social and behavioral science data.

The development of an increasing number of these data centers is an important innovation for the social and behavioral scientists, in that they provide current data necessary for a number of research applications. These centers contain hard core data, such as census information and labor statistics, that are important not only on the national level, but on a local and regional level, as well. Other well-known types of data being incorporated into these centers are welfare statistics, crime statistics, land-use data, transportation information, legal decision, legislative and electorial voting records, and survey data of all types.

The significance of having these kinds of data more readily accessible can be seen in their economic, technological, and social applications. From the economic standpoint, they are indispensible in the assessment of the health of the country, states, regions, and local administrative units. For each of these units, economic data provide the basis for taxation and budgetary allocations. From a social science standpoint, they are necessary in the administration and management of all levels of government, and most importantly, they are the indispensable ingredient of representative democracy.

These data also have broad application in technological areas. The resolution of problems such as air, water, and noise pollution are as dependent on the application of social and behavioral science data as they are on technological and scientific data. Regional balance in the location of industry, the health of our cities, and transportation problems also require for their solution factual data collected by the social and behavioral scientist.

These applications indicate the significance of maintaining a wide variety of accurate data at the disposal of the social and behavioral scientist. Not only do such data resources allow the social scientist, in cooperation with the physical scientist, to find solutions to problems that have social and technical aspects to them, but they also allow for the resolution of basic social and political problems. These data also contribute to a greater understanding of the nature of our social and political institutions. Therefore, in a very direct sense, data pertaining to the social and behavioral sciences have a more immediate impact on the daily lives of the individual citizen than the supporting data in the physical sciences. This is due to the fact that each individual interacts with his social and political environment every day, whether he is applying for a license or voting for a candidate for political office.

The foregoing comments indicate the type and scope of data important to research and progress in the social and behavioral sciences. Simply, the data are about people and their environment. The first requirement of the social sciences, therefore, is to continue to build strong data bases that contain accurate data in a more accessible form than was available in the past and more comprehensive data for purposes of comparison.

Relationship between Data Efforts. For the social sciences, three types of relationships among data efforts exist: the relationship between the data efforts in the social sciences and those in the physical sciences, the relationship among the data efforts of each discipline within the social and behavioral sciences, and the relationship among the data efforts within the same discipline. Little interaction exists between data efforts in the physical sciences and those in the social sciences. Two undesirable drawbacks arise from this lack of communication:

- The first is that few, if any, of the various disciplines can operate in isolation from one another. Each field requires information and knowledge about the progress and direction of the others. Basic and applied science interact, physical and social science interact, and the goal of all disciplines is for increased knowledge about ourselves, our actions, and our surroundings. Solutions to many of the most pressing problems, therefore, require an exchange of data and/or information between these two broad fields of inquiry.

- The second aspect concerns the nature of data itself. Data are the principal commodity in every discipline. Therefore, greater interchange is required, not only with regard to techniques in establishing data centers, but in the exchange and accessibility of data contained in each center. For example, a social scientist who makes use of the social science data archives might well make extended use of information contained in a number of urban data centers. The difficulty in doing so lies in the fact

> that urban data centers are not structured to accommodate research inquiries requiring such data. For those data centers, the purposes of which are similar among the various fields of social science (e.g., scholarly research), the problems of interchange are less formidable, but formatting, costs, and language problems remain barriers to be overcome.

Another relationship that exists between data efforts in similar fields of social science is that a great deal of work is being done to coordinate these efforts through societies, associations, and coordinating councils that are established precisely for this purpose. The problems and achievements concerning this relationship will be dealt with in greater detail in a following section of this report.

A relationship of another kind exists among the data efforts in the social sciences. Hierarchical in nature, this relationship is between federal, state, and local data systems in a particular field. A good example of this kind of relationship can be seen in the area of law enforcement. Local law enforcement data systems depend on and interact with state and federal data centers. Compatibility between these systems is vitally important if effective law enforcement is to be achieved. The case is the same for other areas where the three levels of government are engaged in the same field. In addition to law enforcement, welfare administration and urban renewal are two other examples of areas of mutual interest in the same data. The need for cooperation among the three levels of government to achieve compatibility and efficiency and to minimize redundance in the operation of these data systems is very important.

2. Characteristics of Social and Behavioral Science Data

The characterization of social and behavioral science data has not reached the sophistication achieved in the physical sciences. The categorical terms basic, developmental, and applied have little meaning in the fields of social and behavioral science. Possibly, social science data could be categorized as basic or applied, but to classify the data to such arbitrary categories accomplishes little when the field itself does not make these distinctions. Other categories common to the physical sciences (e.g., degree of refinement, evaluated, degree of accuracy, etc.) are equally as cumbersome when applied to the behavioral sciences. Social and behavioral scientists, however, use the categorization raw data in reference to census records and sample survey. The term clean data is also used in referring to data that are free of errors and clear of ambiguity.

Certain other characteristics that have a parallel in the physical sciences can be applied to social and behavioral sciences. The characteristic of discipline orientation applies to both types of fields. However, almost all social and behavioral science data are discipline-oriented, while data in the physical sciences are more frequently mission-oriented, as in the space and atomic energy programs of the federal government. The characteristic of obsolescence is also applicable to both fields, but, unlike physical science data, social and behavioral science data do not become obsolete very rapidly. In most cases, social and behavioral science data can be used for historical and secondary analysis in the future. Population statistics is an obvious example of data that retain their value, although their usage decreases over a period of time.

In the development of data centers, the volume of data is an important characteristic to be considered. To date, there have been no studies in the social and behavioral sciences that have attempted to ascertain the actual or potential volume of data for inclusion in a single or series of data centers in the social or behavioral sciences. It follows that growth rates cannot be assessed without an estimated volume of existing data within the social and behavioral sciences. So far, the only measure regarding the volume of data in the behavioral sciences is an inventory that will list the number of sample surveys held by the social science data centers. This inventory is presently being prepared by the Council for Social Science Data Archives. The inventory, however, does not provide an easy means of comparison with data collections in the physical sciences.

3. Data Flow - Generators and Users

Social science data are generated in a number of ways. Census data, required by law, are generated by periodically counting the individuals within specified geographic areas. Sample survey, constituting a large segment of social science data, are generated by use of various sampling techniques. Other data are generated by simple tabulations under governmental auspices. These methods constitute the major ways of generating data in the social and behavioral sciences.

The users vary, depending on the type of data of interest to them. Primarily, the users consist of research social scientists, government administrators, sociologists, social workers, and economists. Institutionally, the users include governments, industry, and universities. Members of the ICPA constitute a special group of users on the

university campuses across the country, as mentioned in subsequent paragraphs. Few user studies have been made in the social and behavioral sciences. The outstanding exception is the work of Dr. William Garvey who has done several studies in the field of psychology.

Data Efforts. As of this date, there is only one coordinating body for social science data centers. Established in 1962, the Council for Social Science Data Archives (CSSDA) is a planning, policy-making, and information-disseminating group for coordinating and publicizing the activities of a confederation of social science data archives in the United States. Its most basic principles are that machine-readabl data and supporting documentation useful to the social science community should be readily accessible, at minimum cost to scholars, and be rediffusible to archives and individuals.

The Council performs many functions which the individual archives are not in a position to perform in an effective manner for themselves. It acts as an intermediary between the archive or individual researcher to obtain data from the large suppliers. It facilitates exchanges between domestic and foreign archives, and it seeks to uncover new data sources to be shared by the archives. The Council attempts to identify gaps in coverage, stimulate the development of new archives, or encourage existing archives to expand their coverage. The Council acts as a referral center, directing particular persons or groups to the archive containing data of interest to them. To perform this function effectively, the Council assumes the responsibility of maintaining an up-to-date inventory of the accessible data in each archive. A directory of data archives and a short summary of their contents have been developed to enhance this process. This was done in collaboration with the European Federation of Social Science Data Archives, which is preparing a comparable directory. A detailed inventory of data contained in the centers is now in progress. Not to be overlooked is the job of the Council to set standards among the archives concerning input, formats, and publication of data.

Other functions of the Council are communicative in nature. Conferences, special meetings, and newsletters fall in this category. Exploratory development is still another function of the Council. Under the Council's direction, efforts to establish an experimental telecommunications link between several data archives are being conducted. Other exploratory efforts concern computer development needs of both hardware and software including compatibility problems among systems.

Another formal data effort is the University Consortium for Political Research, a partnership between the Survey Research Center of the University of Michigan and some eighty universities, colleges, and non-profit organizations in the United States and abroad. When it was established in 1962, the Consortium's main objective was to make the data resources of the Survey Research Center available to individuals located at other institutions. The main purpose of the Consortium today is to create an archive of multi-purpose data that will serve a variety of research and training needs and to develop computer-oriented systems of data management and information retrieval designed to maximize the utility of data archives for the individual scholar. The Survey Research Center has one of the largest collections of survey data pertaining to American national elections. Through the Consortium, the survey data of the Center are made available to member universities for research and teaching. The annual membership fee is either $1,500 or $2,500, depending on the type of services desired. The fee constitutes approximately 30% of the Consortium's operating budget. The Survey Research Center is also a member of the Council of Social Science Data Archives.

Social and Behavioral Science Data Centers. There are today approximately twenty-five formally organized social and behavioral science data centers in the United States. In the main, these data centers are located on major university campuses, although two of them are important parts of data operations of the Federal Government, as mentioned later. Most of them are members of the Council of Social Science Data Archives.

Like data centers in the physical sciences, the main purpose of the social and behavioral science data centers is to collect, preserve, and disseminate data to aid and foster research. This is achieved by providing greater accessibility to the data through formal channels. By so doing, far greater use can be made of the data through secondary analysis of surveys.

The contents of these data centers consist of voting and referendum records, census and labor statistics, biographical data, and survey records. These collections are in some cases national in scope, while in others, the coverage is limited to a state or a particular locality. Still others are concerned about a particular area of the world, such as Latin America or Asia.

While the interests and composition of these social and behavioral science data centers may vary, they constitute the most significant efforts to date in developing a viable data base for the social and behavioral sciences.

Of all the formal data efforts concerning science and technology, the development of urban data centers is among the most nationally significant. This is not because such developments advance science in any major way, but because these developments make possible a far more extensive application of technology to impo tant socio-technical problems. Located in our major cities and counties, these data centers are automating many of their functions and corresponding records for the expressed purpose of providing more efficient local government. These efforts include the automation of numerical records (such as welfare, crime, and school statistics), budgets, record keeping procedures, and methods of reporting. Consequently, an urban data center's immediate practical utility is found in better public administration and management of city government, and for closer scrutiny of the social problems confronting our cities.

The impact of the urban data centers is far more extensive than that of efficient local government. For scientists and engineers, the development of these data centers is vital, inasmuch as the centers contain a type of information indispensible to the work of civil engineers, transportation engineers, public health officials, and many other specialists concerned with technological problems that are inhibiting urban progress and vitality. For industry, they are becoming an important source of information concerning product markets, the type and availability of labor forces, and the desirability of plant locations. In some instances, industry has been the sponsor for the development of such data centers.

These data centers promise to be of great value for scholarly research. The social or behavioral scientist could enhance his research materially by using the data contained in these centers, instead of generating his own data or deriving it from other sources tnat did not have his objectives specifically in mind. For example, a comparison of the changing composition of welfare roles, and comparative studies of crime and crime prevention in different cities, are types of research that could be undertaken more profitably with data made available through urban centers. So far, these potential new sources of accurate, up-to-date information have not been utilized by the behavioral scientists for research due to the configuration of these urban data centers as management information systems.

Basically, two types of data are included in these centers: data on citizens, such as welfare records, police records, public health records, etc.; and data concerning the physical aspects of the city, such as highways, modes of transportation, housing, recreational and educational facilities, etc. As management information systems, these centers provide periodic reports to heads of departments and city managers. Increasingly, however, these data centers are developing the capacity to provide the city officials the means to interrogate the system for specific information on an on-line basis.

The reasons given for the development of urban data centers are numerous. The following ones have been taken from published articles discussing various centers in operation or under development. In terms of the routine operations associated with urban affairs, the reasons for the development of urban data centers are:

- To provide a means to handle a large and growing volume of routine paper work;
- To reduce duplication in the collection, storage, and processing of data; and
- To provide greater access to the information in terms of speed, use, and flexibility.

These practical reasons for the development of urban data centers are based on more fundamental reasons, which pertain to managerial operations of urban affairs:

- To provide better information for better decision making;
- To facilitate better urban planning (e.g., physical, social, economic, and fiscal);
- To provide greater capability for managing, controlling, and evaluating the myriad of local government programs within their jurisdiction;
- To provide for more effective control over expenditures;

- To measure operating achievements against planned goals; and

- To allow for the application of the most effective management methods that are available today.

These fundamental reasons naturally lead to the more generalized and far-reaching purposes for the development of urban data centers. These ultimate, socio-political goals are:

- To provide for improvement in the processes of local government;

- To provide for greater cooperation and coordination between local government and local business and industry; and

- To meet the demands of reporting placed on a city by state and federal governments.

The census of data efforts contained in Part C of this volume does not provide an enumeration of urban data centers which are now in operation or planned because of two reasons. First, these data centers do not have as their primary objective the development of a data resource. Their primary goal is not to provide better support for social or behavioral science research, but to provide better support for operations of governmental administration and management. Secondly, little data are available concerning the development and operation of these centers as information sources pertinent in the context of our study. As internal developments within the governmental organization, they are considered an operational data processing efforts and not data resources. Therefore, literature concerning these centers is similar to internal progress reports, as opposed to reports about progress within the field of information transfer. A comprehensive, or even a preliminary, survey is difficult to achieve. However, the following list is provided to

indicate the significance of this development on a national scale. This is only a partial listing of data efforts at the local level that have been reported in the literature or have been contacted or visited during the execution of this contract. These centers are either in operation or are in the planning stage:

Metropolitan Data Center Project
Tulsa Metropolitan Area Planning Commission
Tulsa, Oklahoma

Metropolitan Police Department
Integrated Information System
Washington, D.C.

San Diego Metropolitan Data Bank
Public Affairs Research Institute
San Diego State College
San Diego, California

The San Fernando Valley Reference Book
Center for Urban Studies
San Fernando Valley State College
Northridge, California

Phoenix Law Enforcement Assistance
Development Study (Leads)
City of Phoenix Police Department
Phoenix, Arizona

Regional Management Information Project
Metropolitan Washington Council of Governments
Washington, D.C.

Social Service Information System
Michigan State Department of Social Service
Technology Planning Center
Ann Arbor, Michigan

Basic Land Economic Data for Selected Areas
of Northern Wisconsin
University of Wisconsin
Madison, Wisconsin

The Detroit Social Data Bank
Detroit, Michigan

The Urban Data Center
University of Cincinnati
Cincinnati, Ohio

Alameda County Data Bank
Alameda, California

San Francisco Data Center
San Francisco, California

Urbandoc
New York, New York

South Gate Municipal Management Information System (SOGAMMIS)
University of Southern California
Los Angeles, California

Dade County Data System
Dade County, Florida

Santa Clara County Planning Department
Santa Clara County, California

Tri-State Transportation Study Commission
New York;
Alexandria, Virginia

Bay Area Transportation Study Commission
San Francisco, California

Boston Regional Project
Boston, Massachusetts

New Haven Data Center
New Haven, Connecticut

Perhaps the oldest formally organized data effort in the country is the Census Bureau, which has been collecting data by constitutional mandate almost since the Federal Government began. The Constitution states: "The actual enumeration shall be made within three years after the first meeting of the Congress of the United States, and within every subsequent term of ten years, in such manner as they shall by law direct." It might be well to recall that the original purpose of the census was for the apportionment of representatives and for the direct taxes among the states.

Today, the Bureau of the Census is required by law to gather statistics on population, housing, construction, agriculture, manufacturing, mineral industries, business, transportation, governments, foreign trade, and shipping. These data are collected for the purposes of providing the government, the public, and cooperating groups statistics and related services in the demographic and economic fields. All of the data are available in aggregate form or other statistical measures (e.g., ratio), with the only restriction being the protection of the confidential nature of census returns.

The Bureau recognizes that social and behavioral scientists have an increasing need for its data. In order to make it more accessible to this research community, the Bureau has investigated new methods of presentation and is also preparing magnetic tapes of its aggregated data for direct processing and analysis by social and behavioral scientists. By 1970, it is projected that the Bureau will materially assist urban data centers, and perhaps stimulate the development of new centers by making available to them the census data in machine-readable form in the specified format requested by the city or county.

The only other data effort in the Federal Government of major interest to social and behavioral scientists is the Bureau of Labor Statistics. Like the Census Bureau, the purpose of the Bureau of Labor Statistics (BLS) is to provide the government and various sectors of the economy current economic indicators, such as figures on employment and unemployment, consumer prices, and industrial production. At present, six readily accessible, machine-readable files are maintained by the Bureau:

- Survey of industry labor turnover;
- Survey of scientific and technical personnel in industry;

- Survey of industry employment, worker earnings, and hours;
- Survey of industry employment payroll and hours;
- Estimates of labor force characteristics from current population survey; and
- Survey of consumer expenditures.

The Bureau's membership in the Council of Social Science Data Archives indicates its interest in making the data collection more accessible to the research community of social and behavioral scientists.

4. Principal Problems

The development of social and behavioral science data centers is a major step forward in the conduct of research in the social sciences. These efforts should be supported and expanded by creating new archives and making those already existing more comprehensive. The main hurdle to be crossed in doing so is that of financial support. In this regard, the support of private industry should be actively sought because of the potential benefits of various types of aggregate data that could be useful to them.

On a national scale, there is a lack of communication and knowledge about the coordination of data activities of other scientific and technical communities. The problem created by this situation is the loss of valuable experience gained by others in understanding and solving similar data coordination and structuring problems. The experience received in solving data-handling problems in one area should be widely shared with those in other subject areas. This problem, to a large degree, could be overcome by holding a conference composed of members of the various coordinating bodies of data activities that oversee a number of disciplines. The planning and coordinating bodies of such disciplines as engineering, medicine, pharmacology, oceanography, chemistry, and the social sciences as

participants could share their experiences concerning such subjects as networking, compatibility, structuring, hardware and software applications, and user requirements. The holding of such a conference would create a broader interest in the need for and development of national data-handling systems and correlation of social science with other data.

In the social and behavioral sciences, there is the problem of maintaining an inventory of data resources, first in the United States and second, internationally. Important strides in this direction are now being made. Over and above these inventory efforts is the need for discovering and exploiting other sources of data of wide utility to social and behavioral science research. As stated earlier, urban data centers constitute a potential and valuable resource for social and behavioral science research. Efforts should be undertaken to determine how these resources can be made available to the research community. Other sources such as corporate management information systems could also provide source data of value to social science research.

Another problem of major importance is creating among the research community an awareness of the social and behavioral science data center's resources. This is basically an educational problem that should be resolved within the university curriculum. Not only should a means be provided to learn of the existence of information resources, but instruction in the use of these new and evolving resources should be given. By providing an awareness of and instruction in the use of these data resources at the university level, the utility of the data centers will increase, thereby giving the centers vital information concerning their community of users, the types of information requested, and added knowledge regarding data management and structuring of the data store.

No comprehensive assessment has been made of what constitutes social and behavioral science data. A knowledge of which kinds of data are utilized by researchers, what kinds of data are needed, and what kinds of data do not now exist would be useful. It could then be determined what data should go into social science data centers. Moreover, priorities could be set for their orderly development. Such an assessment could help to determine the condition of the existing data regarding its accuracy and completeness.

Perhaps the most controversial issue concerning data efforts in the social sciences centers around the problem of the right to privacy. Safeguards will have to be developed, not only for the individual, but for communities, as well. The safeguards must be technical; i. e., built into the data system itself, and legal. The problem of the right to privacy, although it now centers on the dangers inherent in large data banks, is a much broader issue that must be viewed within the context of the entire technological revolution.

Selected Bibliography

1. "Do We Need Social Observatories?" Saturday Review, August 5, 1967, pp. 49-52.

2. Ralph Bisco, "Social Science Data Archives: A Review of Developments", The American Political Science Review, Vol. LX, No. 1, March 1966.

I. Environmental and Geosciences

1. Introduction

The physical environment, in the most basic sense, is determined by the land, water, and air masses and their mutual interactions. The study of the environmental data, therefore, involves examination of the geosciences, meteorology, and climatology. Two difficulties inherent to study of environmental data are the complex relationships between associated disciplines and the scope of the required analysis. First, oceanography, due to its unique size and role, is treated separately in the section to follow. Secondly, certain environmental phenomena are involved in all fields of science, engineering, economics, and social/cultural-oriented activities, and these are beyond the scope of this study. Therefore, this discussion is limited to the data activities of the sciences dealing with the measurement and prediction of environmental phenomena. The gross interrelationships of the environmental geosciences and their applications are shown in Table II-I-1.

Meteorology, the science of the atmosphere, has several subdisciplines which result from a need to divide data generating efforts according to altitude, time, and relative size of physical and chemical phenomena to be measured. Weather forecasting is an attempt to predict atmospheric behavior; it deals with the stratosphere to sea level short term circulators on a gross scale. Climatology, by contrast, is the long term historical variations of weather patterns about any given locale. Aeronomy deals with the chemistry and composition of the atmosphere; at lower altitudes, it is concerned with aerosols and pollutants, while at the mesosphere to ionosphere levels, it is concerned with energy transfer mechanisms between the sun and the earth, as well as the role of the various constituents in this process.

Geology, the science of the Earth's structure and behavior, is also a conglomorate of intertwined disciplines. Geophysics is concerned with the heat flows, seismic stresses and disturbances, magnetism and gravitation. Geodesy is the measurement of the earth's surface features, as well as their relative distances; the culmination of geodetic efforts is the determination of the size and shape of the earth. Physical geology is the study of faults, folds, weathering effects, glaciers, volcanic effects, and evolutionary processes which explain the appearance and structure of the surface. Minerology pursues the relationships of crustal features and deposits, chemical and physical processes. Geochemistry concerns itself with the composition of corings, diggings, and surface material.

TABLE II-14. UTILITY OF ENVIRONMENTAL AND GEOSCIENCE DATA

FUNCTION OR RELATED DISCIPLINE	Geology: Mineralogy	Geology: Petrology	Geology: Economic Geology	Geology: Stratigraphy	Geology: Cartography	Geophysics: Geodesy	Geophysics: Geothermometry	Geophysics: Seismology	Geophysics: Tectonophysics	Geophysics: Geomagnetics	Geophysics: Electrics	Geophysics: Ground Water	Geophysics: Limnology	Geochemistry: Organic	Geochemistry: Inorganic	Geochemistry: Isotopic	Aeronomy: Ion Sounding	Aeronomy: Airglow	Aeronomy: Aurora	Aeronomy: Geomagnetics	Aeronomy: Solar Flares	Aeronomy: Density	Aeronomy: Composition	Meteorology: Surveillance	Meteorology: Forecasting	Meteorology: Climatology	Meteorology: Aeronomy
Geochronology		x		x	x		x			x	x	x	x	x	x	x										x	
Geocosmogony		x			x		x	x						x		x										x	x
Geography	x	x	x	x	x		x	x				x	x	x	x												
Planetary Circulation																								x		x	x
Clear Air Turbulance																								x			
Space and Planetary Exploration	x					x	x	x		x				x	x	x		x	x	x	x			x		x	x
Exploration/Prospecting	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x										x	x
Construction	x		x	x	x		x	x				x	x	x										x	x	x	
Agriculture			x		x									x	x	x								x	x	x	
Communications																	x		x		x						
Transportation	x		x	x	x															x	x			x	x	x	
Mining, Drilling	x	x	x	x	x							x	x	x	x	x										x	
Manufacturir			x		x			x				x	x											x	x	x	
Health and Welfare			x		x			x				x	x	x	x									x	x	x	
Logistics			x	x	x							x	x	x	x									x	x	x	x
Strategic Warfare			x		x	x										x	x			x	x			x	x	x	x
Tactical Warfare			x		x							x	x	x	x	x	x							x	x	x	x
Communications					x	x											x										x
Weapon System Development			x		x									x	x	x	x			x				x	x	x	x
Control and Intelligence	x		x	x	x	x		x				x	x	x	x		x	x	x	x	x	x	x	x	x	x	x

Hydrology is an interface between meteorology and geology; it is concerned with transport, compositional changes, and storage of water from the atmosphere back to the sea. Geography integrates geological, economic, cultural, botanical, zoological, and climatological data for any given area of the earth. Cartography, a precision mapping of the earth with periodic updating to reflect natural and man made changes, supplies the base line for geography's integration of other geological efforts. Paleontology deals with the study of ancient life and their environs as described by fossic structures.

The total effort of the environmental and geosciences is difficult to define; however, there are associated bits and pieces of information available that provide a partial view of the field's scope. For instance, the International Geophysical Year (IGY) involved some 66,000 scientists from 66 different countries at a cost of about one billion dollars. The total Federal expenditures in earth science were about 540 million dollars in fiscal '63 and 600 million in fiscal '64. The pre- and post-IGY investment in the environmental sciences is providing us with knowledge that has the following utility:

- Locating, appraising, and conserving natural resources;
- Forecasting weather and modifying climate and weather;
- Reducing damage from violent, self-induced perturbations of the earth - hurricanes, tsunamis, earthquakes and volcanic explosions, all of which are destructive to man's life and his artifacts;
- Preventing and overcoming the effects of pollution from man's advanced activities;
- Designing, testing, and using military weapons and predicting weapons' effects;
- Providing knowledge for improved long distance communications;

- Increasing the economy and efficiency of air and oceanic transportation;
- Developing optimum patterns of land use; and
- Designing engineering structure for the improvement of man's environment.

Scientific and technical involvement in the meteorological field is best measured by the associated Federal budget. In 1967, the estimated total Federal expenditure was about 368 million dollars, of which 183 was funded by the Department of Defense; 110 million dollars was spent by the Department of Commerce, and 92 million was spent by all other departments. Outside of the Federal government, the largest expenditures on meteorological services involve the airlines, insurance, and communication media. An estimate of these expenditures would be in the 3 to 7 million dollar bracket. Of the 368 million spent by the Federal government, 149 was spent on observation, 50 on analysis and forecasts, 53 on communications, 46 on dissemination, and 68 on general support.

An index of the importance and magnitude of the hydrological data activity is an assessment of the activities of the U.S. Geological Survey (USGS) within its Office of Water Data Coordination. The USGS has approximately 10 thousand surface stations and takes data from over 500,000 ground water wells. Sampling time varies from monthly to yearly, depending on the amount of data activity in the water flow in question. The USGS is concerned with such data as flow and water table level, whereas water purity and pollution are handled by state agencies. Since the primary use of the data is for both long and short term trend measurement, its obsolescence is minimal. Typical users include Federal, state, and local governments, as well as private engineering consulting firms. For local trends, the best index of water data availability would be with local government or engineering consulting firms, the latter of which consider their analysis and sampling somewhat in a proprietary vein. A future source of hydrological data will be the satellite. Remote infrared sensing spots fresh water runoff into the sea and monitors snow cover, as well. These data will be qualitative in nature; i.e., mapping, but data processing technology will permit the conversion to tabular formats. In terms of flow, about 3/4 of all rainfall (or snowfall) flows as ground water. Precipitation and glacial melt is the source of water, although volcanic activity will also add to the water balance. The total water budget is affected by long term

climatic variations; hence, the scientific community is also interested in ice formation cycles, oceanic levels. In summary, the volume and national significance of hydrological data are difficult to assess because of the diversity and relatively unstructured nature of the data gathering activity.

The level of data generating activity in the geosciences is best measured within the context of Federal expenditures in the environmental sciences. For instance, Federal expenditures in environmental sciences for FY'63 totaled 540 million, while FY'64 totaled 600 million. Federal expenditures in the solid earth sciences for FY'67 totaled 145 million. This figure includes 12 million for geochemistry, 66 million for geophysics, and 25 million for mineral technology. The commercial investment is quite high. The top 20 petroleum companies are reported to spend 500 million per year in collecting exploration data and another 1 billion in data processing, analysis, and storage.

2. Data Characteristics

In characterizing the data and data activity for the environmental and geosciences, the approach used is to analyze meteorological, hydrological, and geoscience data, in that order.

a. Meteorological Data. For the purpose of this study, meteorology is defined as the study of the planetary circulation, the prediction oriented effort of weather forecasting, upper atmosphere studies of the aeronomy field, weather or climate modification, small scale meteorological effects such as cloud physics, clear air turbulance, sferics, tornadoes, and air pollution. The scientific and technical efforts supporting these fields include hydrodynamics, heat balance, or thermodynamics, molecular physics and chemistry, mathematics, communications and system engineering, instrumentation, and sensor platform technology. Since the atmosphere interacts with land and sea masses, as well as man's cultural institutions, these facets of meteorology must also be considered.

The key distinctions in meteorological data are precision, quantity, accuracy, the means in which they have been gathered, the spatial domain of the data (altitude and horizontal planes), the frequency of observation, and finally, the amount of data processing

and personal interpretation they receive. There are eight functional areas of meteorology:

- Weather Surveillance, wherein the required data vary but include temperature, pressure, humidity, wind vector, cloud patterns via satellite; and ground radar.

- Weather Forecasting, where data are the same as for weather surveillance, but emphasis is on quantitative data over wider area coverage for extended forecasts. Global forecasts on the two-week level also require oceanographic parameters;

- Tropical and Planetary Research, wherein the key weather variables are temperature, humidity, wind vector, parameter gradients, pressure over a finer grid than is the case with forecasting, ocean temperatures, and current flow;

- Atmospheric Chemistry and Pollution Surveillance, which involves aerosol distribution and composition, radiation buildup and dispersal, circulation behavior;

- Clear Air Turbulance Research, which involves wind shear and vector, radar returns with statistical interpretation, infrared and microwave radiometer readings, shock spectrums as a function of atmospheric density, and aircraft speed;

- Cloud Morphogensis, which includes pressure, electrical charge, aerosol and dust distribution, wind vectors, parameter gradients, temperature, surface interaction;

- Climatology, which requires records of temperature, winds, isobaric plots, rainfall, snowfall, as well as oceanographic data;

- Aeronomy, which involves communications performance data, ion classification and distribution, albedo, temperatures, densities, solar flare responses, heat balance, acoustics, radiation intensities, and airglow.

The data characteristics associated with these areas are summarized in Table II-I-2.

TABLE II-I-2

CHARACTERISTICS OF METEOROLOGICAL DATA

Atmospherics (Meteorology)		Scale	Sensor Platform	Degree of Refinement
Weather Surveillance (rainfall, tropical disturbances, fronts)	Temperature, pressure, humidity, wind vector, precipitation, radar display, satellite cloud photography; qualitative and quantitative inputs related to nephanalytic plots.	Areas peripheral to population, industries, military centers in question.	Balloons, buoys, aircraft, ground stations; satellite merchant ships.	Qualitative data; little required for local use; for global and regional uses, considerable computerizing and hand matching required.
Short term Weather Forecasting (1 to 3 days)				Quantitative data; for local needs, little refinement needed. For regional and global uses, fantastic communicating and computer processing required to produce forecast results.
Extended Weather Forecasting (14 to 28 day) (WWW goal)	Key numerical weather variables from global grid measurements.	Global with large grid distances.	Balloons, buoys, aircraft ground stations; satellite merchant ships.	Capability non-existent, awaiting results of GARP. Considerable computer processing envisioned.
Planetary Circulation Research (GARP)	Fine grid data from selected tropical areas; earth heat balance. Vertical sounding parameters.	Regional with fine grid distances, global for heat balance.	Balloons, buoys, aircraft, ground stations; satellite merchant ships.	Considerable computer processing will be required to derive behavior patterns within framework of existing weather mathematic models.
Cloud Morphegenesis Research	Pressure, humidity, aerosol distribution, sterics, wind vector, surface interaction	Local and regional with varying grid requirements.	Aircraft, balloons, ground stations.	Varies from photographs to tabular descriptions of parameters measured and results.
Pollution Surveillance and Countermeasures	Aerosol distribution and composition; circulation patterns; radiation levels.	Local with fine grid requirements.	Ground stations.	Highly specific in terms of trace contaminant count at measurement point. Sensor readout doesn't require additional processing.
Clear Air Turbulence Research and Forecasting	Radar; infra-red; planetary circulation patterns.	Local and regional; fine to large grid results.	Aircraft, ground stations, balloons	Considerable statistical correlation of results required.
Climatology (100 years)	Isobaric plots, key weather variables.	Regional and local.	Compilation of weather surveillance records.	Less refined than local measurements; considerable averaging required.

TABLE II-I-2 (continued)

Presentation	Volume of Data	Rate of Obsolescence	Economic Value	User	Technical Sophistication
Nephanalysis, satellite cloud photos.	Local and regional; 1 satellite photo every 3 hours. Continuous measurements of all available parameters at local level.	Highly perishable; key indicators are used for climatology.	For the U.S., of millions per year in benefits.	Government, commerce, civil; all segments (military, agriculture, industrial, mining, transportation, leisure).	Low in result form, but quite high in intermediate stages.
Nephanalysis.	Considerable to regional forecast centers; from regional to local, volume limited to one to two nephanalyses per day.	" " "	" "	" " "	" "
Nephanalysis of surface patterns, as well as several altitude flows.	Estimate for 14-day forecasts: 100,000 measurement stations, 510 8 measurements/ station; sampling rate varies from hourly to daily.	High degree of perishability after storage periods varying from 1 month to 5 years.	Estimates of payoff: U.S. - 5 billion/year; global - 15 billion/year.	Worldwide and multi-cultural utilization envisioned.	" "
Tabular, graphical satellite photos.	Same volume, but higher rates than extended global forecasts.	No obsolescence envisioned; studies will be used as measurement standard for future research and forecasting operations.	Payoff in early implementation of extended forecasting.	Scientific community.	Quite high.
Photographic; tabular, graphical.	By definition, limited to local areas and evolution of cloud types. Data quite small, when compared to forecasting.	Obsolescence through continued research and refinement.	Relatively small.	Scientific community.	Quite high.
Tabular and graphical.	Estimate for population centers of 1,000,000: Daily measurement of 5 parameters at 50 stations.	No obsolescence envisioned as long as pollution remains threat to biological and economic well-being.	Relatively small, but has wide-ranging economic implications.	Municipal, county, and state governments, heavy industry and universities.	Quite high.
Tabular and graphical.	Considerable.	No obsolescence until CAT prediction is possible; little need for additional storage envisioned after prediction breakthrough.	Small, but critical for continued airline confidence with public.	Commercial airlines and military operations commands.	Quite high.
Tabular and graphical.	Considerable; proportional to extension of forecast time space regimes.	None.	Difficult to assess; quite valuable for any long-term planning.	Wide segments of the population.	Low.

Weather Surveillance and Forecasting is characterized by extremely large volumes of quantitative and qualitative data. Quantitative data encompass the key weather variables; barometric pressure, temperature, humidity, wind vector as derived by balloon and surface borne thermometers, anemometers and barometers. By the mid to late 1970's, satellites will be providing pressure, thermal, and humidity profiles of the atmosphere. Qualitative data for the most part refer to mapping functions; cloud cover via ground radar, satellite cloud cover photos via visual and infrared sensors, and heat balance horizontal plane maps of thermal emissions. In addition, sea buoys and merchant ships provide information on oceanic conditions.

The degree of refinement varies for qualitative and quantitative data. For weather surveillance, ground radars provide 200 to 300 mile circular radius sweeps of the surrounding countryside. These data are of low bandwidth at C-Band frequencies and are transmitted in real time to central displays where they may receive no more than operator inspection or perhaps be recorded by a Polaroid-type camera. Satellite cloud cover photos, on the other hand, cover about 1,000 mile swaths as they pass over the earth; they provide 10 to 50 mile resolution photos which require varying degrees of data processing. For instance, local readout schemes transfer the image to photo sensitive papers which then are the end of the refinement cycle. Global readout devices from polar orbiting satellites require considerable stitching and matching via computer before a truly global cloud map is obtained. From synchronous orbit, monochrome and color photos are obtained every hour or [illegible] as opposed to the once every 8 to 12 hour readout of polar orbiters. In all cases, the visual presentation is the end product, which then is subject to operator interpretation of weather conditions. All global satellite data are first recorded on magnetic tape before being converted to photo image.

Quantitative data are gathered and transmitted to the Suitland, Maryland ESSA facility; they are fed to computerized atmospheric models which, in turn, provide an x-y plot with alphanumerics; this plot, known as a nephanalysis, is the weather forecast that is provided to the civil community. Presently, no satellites are providing quantitative data for weather surveillance or forecasting.

Weather surveillance and forecasting are both characterized by high volumes of data along with rapid obsolescence with considerable costs involved. For instance, the Federal FY'67 expenditure of 368 million on meteorological services and supporting research includes about 150 million for observation, 50 million for analysis and forecasts, 53 million for communications, and 46 million for dissemination. In terms of data flow for a typical fiscal year, satellites produce over 100,000 usable photos per year. Quantitative information can be estimated; about 250 to 300 ground stations in the U.S. alone provide readings of temperature humidity, wind vector and barometric pressure about four times daily for one to two altitude levels. In addition, nephanalysis and other numerical data are received from cooperating nations on a global basis. While all these data become obsolete rapidly, much of the data are stored for historical purposes which permit climatology studies. Meteorological data for surveillance and forecasting are oriented toward a local, regional, continental, and global basis. No proprietary or ownership considerations are involved (outside of agency rivalries), since 99 percent of funds are supplied by the Federal Government. On an international basis, cooperative agreements are established for exchanging meteorological data, but in some cases, this is a haphazard operation.

Tropical and Planetary Circulation Research. The data required are similar to those for weather surveillance and forecasting, except that the grid distance between measurements is smaller, while the area of coverage is much larger. There is more emphasis on accuracy, since the research objectives are to discover physical processes which will, in turn, permit a choice of instrumentation and data processing for extended global forecasts. The main programs involved in this international effort are the Line Island Experiments, the Barbados Experiments, along with other tropical research under the heading of TROMEX (Tropical Meteorological Experiments) and GLOMEX (Global Meteorological Experiments); both GLOMEX and TROMEX are constituents of GARP (Global Atmospheric Research Program). Typical data requirements for GARP are shown in Tables II-I-3, II-I-4, and II-I-5. They indicate the large amounts of data needed for global circulation studies.

TABLE II-1-3. SYNOPTIC NET DATA REQUIREMENTS FOR GLOBAL ATMOSPHERIC RESEARCH PROGRAM

Quantity	Vertical Resolution	Accuracy	Remarks
Horizontal wind	Surface, plus normal rawinsonde (or horizontal or volume averages) at or about 925, 700, 400, 200, 50 mb.	1 m/sec.	Averaging desirable.
Temperature or virtual temperature	Same	0.5°C	
Humidity	Same but not above 300 mb.	5% relative humidity necessary 1% relative humidity desirable	
Geopotential height vs. pressure	Airplane or horizontal sounding balloon heights, if available	10 m.	May partially substitute for temperatures.
Pressure	Surface	0.3 mb necessary 0.1 mb desirable	May only be necessary on boundaries.
Terrestrial radiation flux divergence	Same as temperature and wind	1°C/day (.2°C/day in vertical average)	24 hour frequency adequate. lower space resolution possible.
Rainfall	Surface	1 mm/day in horizontal average	Satellite or other data used to improve horizontal average.
Sea surface temperature	Surface	.25°C	
or Direct air-sea temperature difference	Surface	.25° C	
Freezing and condensation nuclei and aerosols	As desired	Uncertain	Presumably from airplane or surface stations.
Cirrus cloud cover	As available	Best available	Partly redundant with radiation flux.

TABLE II-I-4. MASO-NET DATA REQUIREMENTS FOR GLOBAL ATMOSPHERIC RESEARCH PROGRAM

Quantity	Vertical Levels	Horizontal Resolution	Frequency	Accuracy	Remarks
Divergent wind	925 mb necessary 650 mb, 200 mb desirable	150 km necessary 50 km desirable	3 hours	.5 m/sec necessary	Area, volume, and/or time averaging desirable
Stream wind	Same as above	150 km	6 hours	1 m/sec	Same as above.
Temperature (or virtual temperature)	Surface and 50 mb intervals to tropopause	150-50 km plus some additional in cloud groups	3 hours	.5°C necessary .1°C desirable 1°C above 400 mb	Same. May be mostly replaced by more vertical resolution in winds
Sea surface temperature or air-sea temperature difference	Surface	150-50 km	3 hours	0.25°C necessary 0.1°C desirable	
Humidity	Same as temperature but not necessary above 300 mb	150-50 km	3 hours	5% relative necessary 1% relative desirable	Averaging desirable.
Horizontal pressure gradient	Surface	150-50 km	3 hours	0.01 mb/100 km desirable	Boundary values may be sufficient.
Terrestrial radiation flux divergence	Same as temperature	150 km	12 hours	1°C/day (.2°C/day in vertical average)	Averaging desirable.
Turbulent heat, moisture, and momentum flux	Moist layer and cloud groups	Uncertain	Uncertain	Best available	May be calculated as residual.
Rainfall	Surface	150-50 km and area average	3 hours	1 mm/day in area average	

TABLE II-1-5

SATELLITE AND RADAR DATA REQUIREMENTS

FOR GLOBAL ATMOSPHERE RESEARCH PROGRAM

Quantity	Horizontal Resolution	Accuracy
ATS Satellite cloud photography	5 km necessary 1 km desirable	Similar to ATS
ATS Satellite window radiation	10 km necessary 1 km desirable	5° C absolute 2° C relative
Horizontal and vertical scanning radar, ground based, continuous operation.	Complete coverage of meso-net	Calibrated for liquid water content

Three classes of data refinement can be envisioned within the scope of planetary circulation research data:

- Ground based sensor (balloons, buoys, aircraft) quantitative data - These data are collected and fed directly (after scaling and adjustments for sampling rates and grid interpolation) into computer simulated models of the atmosphere. The readouts are then checked against actual weather behavior. In this simulation activity, the objective is to define the optimum data network, the best combination of parameters and finally select the math models that appear to have the best tradeoff in forecast time and accuracy versus cost.

- Satellite based quantitative data - These data are characterized by the fact that the satellite provides a vertical probe of the atmosphere (for temperature and gas distribution). These vertical profilings start with infrared and microwave radiometer readings; the spectral-amplitude profiles then require computer processing to provide an altitude versus temperature and gas distribution profile. Once the profiling is accomplished, the data are entered into computer simulation much like the ground based sensor data approach.

- Satellite qualitative data - These data consist of visual band cloud cover photos with resolution varying from 5 to 50 miles at the satellite nadir. Presently, this information is used for differential adjustments to numerical forecasts and hence, only require man's inspection. However, if implemented on an operational basis to yield the maximum information, pattern recognition data processing could be involved. Infrared data mapping is also qualitative; resolutions for nighttime mapping are in the 20 to 50 mile range, while heat budget mapping has spatial resolutions in the 200 to 500 mile range. Heat budget analysis requires computer processing to relate the infrared readings to geographic projections.

The volume of data involved in research is tremendous; furthermore, research data do not have the perishability factor of forecasting data other than the fact that usually, one finds the quantity and quality of the initial attempt data inadequate, hence triggering requirements for more and better data. The volume of these data is difficult to pinpoint, since the research program is still in its infant stage. However, we do know that Tiros, ESSA and Nimbus have provided about 2-million cloud cover photos over the past four years. ATS, the most recent satellite in the meteorological applications family, has probably provided several thousand cloud cover shots. Nimbus B and D are the first of the vertical sounding satellites, but they have yet to fly; their data will be quantitative and should reach astronomical proportions. From the ground, supplementary sensors will be required. This involves about 20,000 stations, with each station supplying two to four parameters at a sampling rate varying from hourly to daily.

Atmospheric chemistry - This data activity documents research and surveillance of natural and man made composition changes of the atmosphere. Efforts are concentrated on measuring sulfur dioxide, nitrogen dioxide, carbon dioxide, carbon monoxide, ozone, hydrogen sulfide, exhaust contamination from industrial wastes, internal combustion engines, aircraft engines and rocket booster exhausts, and nuclear fallout. Aerosols include soot and fly ash content, bacteria, pollen grains, sea spray and meteroid dust.

These data are tabular and graphical in their final form; as a rule, they require laboratory processing of atmospheric samples before tabulation and final documentation occur. The best index of their volume would be counts of applicable journals and articles per year; however, due to the diverse nature of atmospheric chemistry, many journals are involved and no estimate is available. These data do not suffer from obsolescence, since they maintain value for trend analysis. Orientation is primarily scientific, although data used for pollution alerts tend to be engineering-oriented. Air pollution data costs are not readily available; however, air pollution damage is estimated at 11 billion yearly. Undoubtedly, whatever the data costs, the cost/benefit rates must be low.

Cloud Morphogenesis - These data reflect the following activities: tornado studies, rain making and cloud seeding experiments, aerosol catalyst activity in precipitation probability, sferics, cloud pattern recognition, clouds as tracers of tropo and stratospheric wind vectors, cloud interference with heat balance, hailstorm detection. Sensors in this activity include radar, infrared, barometric and thermometric sensors aboard sounding balloons, aenometers, uhf radiometers, and conventional photography. The data are a mix of qualitative and quantitative format with considerable emphasis on translating the data into statistical presentation.

Climatology Data - These data, for the most part, are a compilation of weather surveillance readings; however, since these can only be as good as recorded history permits, there is activity in determining climate trends from other tracers. Typical sources of pre-historic climatology data include tree ring analysis, glacial traces, shoreline variations, exploration of the continental shelf, radioisotope dating of fossils and the extrapolation of their ecological niche, inferences from geological strata, and other paleontological data.

Aeronomy Data - These data reflect studies of the meso and ionosphere; typical parameters of interest are electron density, ion composition, density, spectral filtering, aurora, magnetics, airglow, planetary limb determination, wind studies via sodium releases, temperatures and chemical species. In addition, considerable research is oriented toward radio transmission characteristics of the upper atmosphere for tropscatter communications and over the horizon radar technology. Upper atmospheric data are utilized for horizon tracker design, satellite stabilization techniques, monitoring of rocket booster exhausts and scientific questions dealing with interplanetary particles, as well as the evolution of the atmosphere.

The question of data volume and degree of refinement for aeronomy data must be approached in the same manner that one would examine laboratory data; they are diverse in their instrumentation source and many of the instruments are custom designed for one reading. The majority of aeronomy data come from sounding rockets; as of 1967, about 7,000 rockets have been fired from 25 U.S. launching stations scattered in the Western Hemisphere. (The world-wide total of sounding rocket sites is about 150.) Since each rocket carries several instruments which provide about 5 to 10 minutes of reading,

one is at loss to define the exact amount of data spewed forth. The data recovery process is either telemetry with magnetic tape recording or on-board recording with parachute recovery. In either case, data processing is involved to strip away the telemetry format to provide the sensor reading. Further data processing may be rewired to achieve the desired data form, such as histograms or geometrically oriented readings. Since the sounding rocket data is scientific, there is no proprietary consideration involved. However, a fair amount of sounding rocket data supports military test and development exercises, and some security may be involved.

In terms of value of these data, the best available index is the cost of one sounding rocket shot. A NIKE-Apache shot runs about 30 to 40 thousand dollars. This price includes the cost of the vehicle and instrumentation package, but not the ground support equipment (which may only require minor modifications for each instrument package). On the other hand, an Aerobee rocket may cost 100 to 150 thousand. And, while no distribution of rocket types is available, it is clear that the 7,600 rockets launched to date represent a considerable investment in meteorological sounding.

Another source of upper atmosphere data deals with ionospheric effects on telecommunications. There is a considerable amount of data dealing with radio frequency sounding. The raw data are recorded on magnetic tape or oscilloscope photos. From there, they are analyzed and presented in the form of histograms, tables, maps and descriptive reports. These data are generated by the scientific, commercial, communications, and military communities.

The most recent voluminous contributor to aeronomy and other upper atmosphere phenomena is the satellite. Here, several objectives are pursued. These include studies of the aurora, radiation belts, density and drag phenomena, airglow albedo planetary limb definition, solar particle emission and geomagnetic field variations. The magnitude and value of these data are not readily obtained; however, a reasonable estimate can be made. Since the IGY, at least 100 U.S. satellites have orbited in the near earth regime which is of interest to the aeronomist and upper atmosphere physicist. Assuming launch vehicle costs of about 1 million per satellite and satellite costs varying between 0.5 and 2.5 million for the mini-satellite classes (observatory class satellites are not included in this estimate), one can envision

total costs of 150 to 200 million dollars for obtaining the data. Again, it is difficult to survey the total volume of data involved, but estimates are again applicable. The on-board data collectors are in the low to medium rate range (a quantitative figure again is elusive), but the data readouts are fairly well determined by the time bandwidth constraints of the telemetry and tracking networks and the altitude of the satellite (which determines data dump time to the ground station). Due to the altitude limits of the low earth orbiters, the data dump time varies from 8 to 15 minutes. Thus, one satellite will transmit scientific and housekeeping data over the standard 10 MH_z bandwidth telemetry channel in the dump time once every pass. Thus, we have 10 megabitsper sec. dumptime per pass which must be recorded on magnetic tape. From these figures, the total data output of 100 satellites whose mean life is in the 6 month to 1 year range is simply a matter of arithmetic estimation which, hopefully, has compensating, rather than cumulative errors.

b.. Hydrological data. These data define the flow, chemical and biological composition of water. By this definition, the user is concerned with the sources and sinks of water flow; ideally, it is possible to trace such a flow from rain and snow in the highlands down its course to the river estuaries. Consequently, the following parameters are of interest: rainfall quantities and runoff rates, snowfall and melting rates, storage within man made and natural lakes or reservoirs, river depth and flow rates, evaporation and ground water flow. Ideally, these parameters would be available as a function of seasonal time along with their averages and anomaly values. This same water accumulates biological excretia. Some filtering and processing may occur along the way to improve its quality. As the water approaches the user, whether it be agricultural, industrial, municipal, or consumer, an economic factor is added to the flow variables, thus adding cost data to the flow and content information. In addition, what was once a problem resolved by nature (the flow) is now a question of national policy, state and regional rights; hence, the data catalogue on water is joined by additional information - that of legal restrictions and interpretations.

As in other fields of endeavor, the data used in this entire process are not a continuum; estimates often are used in lieu of measurements, and variations about the norms are not fully known. However, flow/composition parameters are highly documented in areas of special interest, particularly those associated with man's activities. It is possible to assume that these islands of data, whatever their utilization, are

obviously satisfying their needs or otherwise, they would not have evolved to their present configuration. However, in view of two certain pressures - that of an expanding population, as well as an increase in the economic and aesthetic worth of water - the existing islands of information may not be adequate for the future. This does not imply that more and better data will simply resolve all questions regarding water, nor does it state that data will suffice in lieu of water. It mainly implies that water flow, composition change, utilization, and the questions of economics, aesthetics and legalisms are tied to a common baseline, namely that of hydrological data. Since our society is evolving to more complex levels which can only be supported by rational decision cycles, the need for adequate information is self-evident.

These points are amplified in an NAS-NRC report on water management:

> "Wise solutions to water problems require accurate information about water and the immense diversity of conditions under which it occurs and is used; they call for clarity in judging the value of water and associated resources. These solutions can be reached only when the organization of planning permits balanced consideration of the choices and values involved.
>
> "Information needed is of three kinds: (1) information on the behavior of water and on the ways in which environmental changes affect water as a resource; (2) information on new and more efficient processes of waste treatment, desalting, and water use; and (3) information on user behavior, on the planning and decision processes, and on probable changes in water use as a result of changes in our technology and in our style of life.
>
> "There is much information on how water moves in the hydrologic cycle, and on how to construct dams, canals, and purification works. Less is known of the biological and social effects of such constructions. Much remains to be learned of the way water-use decisions are reached at the various levels of government and in the private sector. We

especially need information that will help increase the number of feasible alternatives and improve water-use decisions."

Looking at the islands of data on water, it is observed that a number of ways exist to classify them. And, while a complete classification is beyond the scope of this report, there are a wide variety of samples to exemplify. Institutions such as local, state, and Federal government all have a different emphasis on data requirements. This is also true of universities, industries, professional societies, non-profit consulting firms, and private consultants. In the most basic sense, the data consist of a quantity measurement at one location which is taken often enough to reflect seasonal variations. From the standpoint of composition, chemical content has historical precedent over other types of content data. This includes salinity, pH factor, sedimentation, temperature. Adding to this chemical data, biological information on limnology is found. Such factors as icthylogical, botanical and microzoological studies document nature's use of water to support life, while man's pollution efforts are noted in thermal, radiological, inorganic and organic changes.

The mix of biological, chemical and meteorological data used in water management varies with application. While all facets of this matrix are far too difficult to describe, several of the following examples point out the diversity of hydrological data requirements versus applications:

- Agriculture - rainfall, soil leeching, surface water chemistry, melting and runoff rates, evaporation rates, irrigation network capacities and costs, river basin flood patterns, well water depths and flow potential;

- Industrial - water flow, inlet and outlet temperatures where water is used for cooling, settling pond characteristics, acid runoffs, water transportation networks (for moving raw materials and finished products), surface and ground water flows affecting foundations;

- Municipal and domestic - water flow, piping corrosion, standtank pressures, water purity, sewerage dilution, swimming facility control, health assessment, rainfall, fire prevention factors;

- State government - flow data for flood control and warning, consumer usage rates, agricultural responses to water cycles, industry recruiting; pollution surveillance standards and data;

- Federal government - chemical composition, flow variables, meteorological trends, economic and legal information, engineering standards;

- Scientific - air/surface interaction such as rain and snowfall and evaporation rates; interface of soil types and water storage characteristics, climatology, ecological studies, erosion and weathering, glacial cycling, trace chemistry, volcanic eruptions which add to the water budget.

The nature of hydrological data varies from a simple well reading to a complex bio-chemical assay; hence, we find the data lends itself to descriptive, tabular and graphic format. Within this framework, we find varying degrees of refinement. A water sample can reflect nothing more than field measurements and analyses taken on a yearly basis, or it can require extensive laboratory processes. In industrial and domesticated regions, increased frequency of measurement is important, as is accuracy of the analysis.

The question of volume, value, and rate of obsolescence for hydrological data simply eludes any accurate reporting. However, enough information is available to indicate the scope of this activity. For instance, about 600 to 700 agencies, including government, industry, private research, engineering, as well as universities, have some aspect of water management and research in their charters. There are figures available

that indicate that the combined efforts of all states and local/regional or industry-oriented organizations are of somewhat the same magnitude as those of the Federal government. For instance:

- The U.S. Geological Survey coordinates data gathering on a national level. Their library (Water Resources Division) is composed of 25,000 books; 252 million items of raw data; 2,000 data compilations; 100,000 well core samples; and 50 analog models of water flow. As for data gathering networks, the available figures indicate that the Water Resources Division maintains 8,200 stream gaging stations; 16,000 ground water wells; 1,600 water quality analysis stations. Another source of information on the scope of all U.S.G.S. data gathering networks indicates some overlap and some conflict with the Water Resources Division figures; specifically, all U.S.G.S. data stations total up to 7 to 10 thousand surface stations and over 500,000 ground water wells.

- Other Department of Interior library holdings further indicate the magnitude of water data: (a) the Bureau of Reclamation maintains 150,000 reports; 600 journals; 18,000 books; 19,000 standards and specifications; 3,200 project reports; 475 data compilations; and 800 research data items. (b) The Office of Saline Water maintains 500 books; 500 journals; 4,000 reports; 2,000 photos; and 3,000 slides.

- From a health standpoint, the Federal holdings of water data are also voluminous. Health, Education and Welfare maintains 10,000 documents on fluoridation of water, while the Army Biological Center maintains 50,000 books, 1,000 periodicals, and 40,000 research documents.

- Other Federal data centers on water resource management include the Army Map Service with over 150,000 maps; ESSA's Environmental Data Service with 135 years worth of meteorological data; and the Weather Bureau's wide range of data compilations, which also include 150 volumes.

On a state and regional level, other interesting figures are available, even though they only provide a sample of what can be found with a very extensive survey.

- Colorado State University - library holdings of 8,000 reports, 3,000 photos, 500 maps and 10,000 sheets of data reflecting 30 to 125 years of observation. The data collecting network consists of 2,500 stations in the Western U.S. and Southwest Canada (stations include ground water, surface water, meteorological and snow survey courses);

- Pacific Northwest Laboratories - holdings of 70,000 books, and 200,000 reports from a data network consisting of 20 meteorological towers, 20 remote telemetering stations, and 500 deep wells in the northwest U.S.;

- Illinois Water Survey - holdings of 10,000 books, 150 journals, and data from a network consisting of 144 steam gaging stations;

- New Hampshire Department of Resources and Economic Development - data from 8,000 drilled wells;

- Maryland Department of Water Resources - holdings of 3,000 reports, 38,000 water quality reports, and 50,000 data compilations from an unknown number of ground and surface stations in the Chesapeake valley basin;

- Pennsylvania - The Department of Forests and Water maintains a network of 350 rain gages and 219 stream gages for flood warning, while the Pennsylvania Department of Health maintains 176 stream gaging stations for pollution monitoring and resource management; and

- The Tennessee Valley Authority - holdings of 3,600,000 data compilations, 2,000 slides, 700,000 microfilms; 18,000 maps; 300,000 engineering drawings; and 1,000,000 punch cards.

These examples indicating the amount and variety of hydrological data provide a good index of the magnitude of volume. The question of obsolescence and the value of these data can only be sheer speculation. Does one consider the value of data, the storage costs, or the station and facility cost, or perhaps, the damages brought on by the lack of these facilities (the cost/benefit ratio argument)? As for obsolescence, one can expect varying degrees. For instance, data that serve an immediate need of assessing flood warnings or pollution content are also valuable for trend data that are needed to locate new sources of pollution or perhaps justify the construction of new dams and locks.

As for the question of proprietary aspects of hydrological data, most of the data are accumulated by government funds; hence, no problem arises. On a local level, however, these data are massaged by private laboratories or engineering consulting firms, at which time proprietary aspects enter the picture. For instance, in any given community, engineering firms that serve a number of small municipalities probably have the best picture of the relationships of ground water levels and sewerage flows.

c. Geological Data. The spectrum of geological data is all encompassing simply because geology draws on all the environmental sciences to reach its conclusions. Assuming a long term goal of perfectly describing the earth, its evolution, and behavior, geology is dependent on all forms of data to further its ends. Some of the types of data and their uses are:

- Geology: surface and subsurface mapping; surface and drill solid samples; erosion, weathering, glaciation and upheaval patterns; outcroppings, regional soil and rock load bearing and stress factors, patterns, volcanic residues and activity and use of any geophysical or geochemical data that describe the features and dynamics of the earth.

- Geophysics: seismic, magnetic, gravitational, thermal and tidal (ocean and lunar perturbations); geodetic measurements of earth bulge and datum plane linkage.

- Cartography: precision mapping of the earth's surface with emphasis on areas of man's activities; this relates closely to geodetics.

- Mineralogy and Pertology: classification of rocks and mineral-bearing solids.

- Geochemistry: pH factors, organic composition of solids, isotopic radiation levels, geochronological dating by isotopes and marine sediments. Thermal and pressure environments and their effective solids crystalline structure.

Surface structures require aerial and satellite photography which provide qualitative data. Local inspections of soil and rock formations are described in tabular and narrative format. Coring samples require storage; they represent unprocessed data. Seismic profiles reveal subsurface structure; these data are recorded in analog format on magnetic tape and strip chart format. Crystallization processes which shed light on the regional and local solids formation require laboratory high temperature and pressure data. Physical deformations of the

surface are described by weathering, glacial and volcanic patterns, as well as uplifting the advance and recession of the ocean levels. And finally, satellites shed light on the distribution of the earth's mass as well as its shape with respect to the geoid. Astronomical data are needed to relate earth wobble and interplanetary tidal effects. Consequently, one is at a loss to describe the total geological effort unless more specific characterizations of the associated geoscience fields are presented. Exemplary information follows on characteristics of seismic, gravitational, magnetic field, cartographic, geographic, geochemical, and geothermal data.

Seismic Data. Two general classes of seismic data exist: measurements of self-induced vibrations from theoretical flows of the earth's core, crustal stresses and lunar/tidal vibrations. The remaining group are those induced by man's activities; these include geological and oceanographic explorations where high explosives supply the perturbations, nuclear explosion detonations and on the microscale, background noise from industrial and cultural activities. Seismic data, since they represent the energy spectrum of earth disturbances, appear in time vs. amplitude plots and time vs. frequency plots.

In terms of refinement, seismic data can vary from direct analog readout (such as the case from a seismometer with a wet pan readout) to large aperture arrays which require computer processing to determine directivity and energy distribution; data processing includes correlation, convolution, spectral densities and digital noise filtering. Geodetic satellites, due to refinements in laser ranging accuracy, are expected to measure earth kinematics in the 1970's. This activity will require translation of ranging data into seismic parameters.

The question of data volume, flow obsolescence and ownership for each class of disturbance must be further qualified by mobility of the sensor. Fixed location small array sensors are oriented towards long wave, large amplitude vibrations and shocks. A large number, perhaps 100 to 150, are maintained by universities throughout the world at an estimated cost of 25 to 30 thousand dollars per year. About one million earthquakes are recorded yearly, of which 150,000 are substantial tremors. More than likely, only major events are preserved in graphical and tabular form. These data do not fall into obsolescence, since they are used for trend and postmortum analyses. The Coast and Geodetic Survey, for instance, is still analyzing data on the 1964 Alaskan earthquake. Other sensors include strainmeters and

tilt meters. Earthquake research in the U.S. is valued at 7.5 million per year; the U.S.G.S. has advocated a 10-year research program costing about 200 million, but this program has not been authorized.

Field seismometers are primarily used for oil exploration and subsurface profiling. Perhaps 300 to 400 such instruments exist; they are used for local sounding as opposed to the global nature of the fixed array seismometers. No figure on volume of data is available, although it could be determined by surveying geophysical exploration departments of petroleum companies. However, costs of oil exploration by the top 20 U.S. firms are estimated at 1.5 billion per year. These data are highly proprietary, with a security approaching that of atomic weapons technology. In terms of obsolescence, old data find continuous use. For instance, soundings in the Persian Gulf have been difficult to shoot, due to reflection from water thermal inversion layers. The old data, however, when subjected to digital noise filtering, revealed petroleum deposit characteristics.

Fixed location, multiple array sensors are used by government agencies concerned with arms control. They have deployed a multiple element seismic array to detect underground and surface atomic explosions. A 625 element array is located in Montana; this array provides a high degree of directional sensing and is sensitive to low noise level tremors. It requires considerable computer processing; its results would provide the scientific community with an excellent tool if it were not for security classification. The cost of this program has been 135 million to date.

Due to inherent increases in position determination accuracy, geodetic satellites are expected to measure earth kinematics. The breakthrough in ranging from 10 meters or 2,000 miles to 1 meter and for the future, 20 centimeters, indicates a large amount of seismic data will come from the space program. The space geodetic program also anticipates placing a laser corner reflector on the moon to determine lunar perturbations effects on earth seismic stresses. Data flow and rates would be on the small order of magnitude as exists with present programs; namely, 200 to 400 laser ranging points per 6 minute sighting intervals. The program is expected to run about 3 to 4 million per year. The data would be available to the worldwide scientific community.

Gravitational and Magnetic Field Data. These data have several common characteristics. First, it can be on a local small scale from surface sensors or it can be from a satellite usage and often, both gravitational and magnetic data are needed to serve the same end. Local gravitational and magnetic anomalies are indicative of mineral or petroleum deposits; mass deflection and electromagnetic sensing coils, respectively, provide the local data. Aircraft surveillance has become increasingly important; an instrument pod with aerodynamic stabilization trails the aircraft to map the area. The data appear as an analog-amplitude vs. space vs. time presentation. These data can be recorded on magnetic tape for computerized transformation to an area map. This information is highly proprietary; no index of its volume (it should be fairly large) or cost is available. This same type of information over the sea is valuable for anti-submarine warfare; some security classification may exist. No cost of ASW gravitational and magnetic anomaly data is available.

Gravity measurements by surface pendulum stations are of great value to the geologist. In Alaska, for instance, 5,000 surface measurements have been taken since 1958; half of the readings were taken at one to two mile intervals. Associated with surface exploration are calibration stations; these points are periodically checked for standard deviations. The Ottawa-Washington calibration range has been calibrated 98 times in 13 years; 12 stations are involved.

From outer space, satellites have contributed large amounts of gross scale geomagnetic and gravitational data. Gravitational variations from space are indicative of relative mass deficiencies of the earth and its shape. Both of these factors are summarized into the zonal and non-zonal harmonics of the Legendre Polynonomial; the polynomial describing the earth shape and resulting gravitational field variation are expressed as "J" coefficients. Considerable radar and optical tracking, along with precision timing, is needed to discover orbital perturbations resulting from the gravity anomalies. Hence, the flow of data is quite large. GEOSI, for instance, after 4,369 orbits, provided 30,886 doppler passes and 1,100 radar passes. The data gathering and triangulation net consists of 185 sites. Perhaps 100,000 sightings per year have been taken so far and considerable work remains to be done. The results of these sightings, after considerable calculations, are simply a table of "J" coefficients and

a map showing average and anomalous gravitational values. Cost estimates are difficult to ascertain; perhaps 200 to 300 engineering and scientific personnel are involved in data reduction and analysis. All satellites and combinations of orbits are observed, even though satellites designed specifically for earth shape measurements have been launched; total costs of geodetic satellites have run about 40 million to date. While there is some military usage of this information it is also available to the scientific community.

For magnetic field data of a large scale, considerable data have come from space probes of the ionosphere and magnetosphere; this application has been purely scientific. Since the sensors are part of multipurpose sensor configurations, it would be difficult to assess this volume and cost. No element of data perishability is involved.

At lower altitudes (100-mile), an attempt will be made to measure local field variations in much the same manner as the aircraft survey. This mineral prospecting can be expected in the '70's; no index of data flow is available to date. These data, incidentally, raise a question of proprietary rights. The government that flies the satellite and secures the data can do so without securing exploration contracts, as would be the case with aircraft. Space lawyers are currently wresting with this data utilization/proprietary question. From the surface, a large number of ground stations provide magnetic data; about 150,000 worldwide magnetic field stations have been surveyed. About 300 to 400 are under periodic inspection.

Cartography, Geography, and Datum Plane Linkage. Precision mapping and geodetic data are somewhat linked; aircraft and satellite mapping describes the horizontal surface features of the earth, while geodesy, via satellites, links precise selected points about the globe within a 10-meter accuracy. Aerial photographic surveys vary in their location, altitude, and degree of resolution. Sensors include wet film camera and side and forward looking mapping radars. Wet film photography requires the conventional chemical development, while radar mapping requires magnetic tape recording prior to transforming the data to images acceptable to the eye. No figures on the quantity of these data are available, but a survey of film manufacturers could provide such. Similarly, radar mapping efforts could also be estimated. Presently, the amount of the world adequately mapped is a nebulous 50 percent. (The qualification for "adequate" is not available.)

Satellites contribute heavily to the flow of cartographic data; presently, this is restricted to the operations of military reconnaisance satellites, but for the future, civil applications from the remote sensor should provide excellent maps of urban areas every 30 days or so, rather than the 5-year cycle that seems to be the rule now. Satellite data are also applicable to forestry and agriculture. Remote data sensing could extend our capability in several existing efforts, providing the data recovery conversion and dissemination problems can be resolved. The applications include government crop compliance checking, crop forecasting, inactive forage range surveys, soil mapping, and disease and pestilence surveillance and detection. Presently, 30 million dollars are being spent annually on forest fire detection and 3 million on crop pestilence/ disease outbreak surveys.

Geochemical and Geothermal Data. The chemistry of the solid earth is becoming an increasingly important class of environmental data. It is characterized less by vast automated collection networks than other forms of environmental data; but on the other hand, there are many more variables to pursue. Consequently, the key format of the data is narrative, tabular, and abstract. Several reasons account for the growth in geochemical and thermal data. They are: 1) a growing concern over the dwindling stocks of natural resources in the face of increased demand; 2) a need to dispense of industrial and radioactive wastes; 3) prediction of earth process behavior and analysis of evolutionary development; 4) harnessing of earth heat for commercial power.

In addition to conventional laboratory equipment, several new instruments are contributing to the pursuit of these data. These include optical emission spectrographs, X-ray fluorescence spectrographs, mass spectrometers, atomic absorption spectrometers, and electron microprobes. At least 30 laboratories have a 50-kilobar pressure capability; there is a trend to high pressure - high temperature studies of the molten to solid state processes.

Analyzing the flow of geochemical data is more difficult than for other classes of environmental data, because there are no vast permanent collection networks, as would be the case with seismology or meteorology. Instead, the laboratory is the collection network, and certain deep wells (or their cores) and surface deposits are the prime source of information. While a survey of laboratories is beyond the scope of this study, some index of their cumulative output is available. There has been a well-defined growth

in mineralogical and geochemical abstracts (which provide access to the data disseminating documents); in 1952, there were 1,300; in 1959, 3,600, and in 1966, 5,600. In new minerals discovered per year, the average is between 50 and 70 per year, since the advent of the newer techniques and instrumentation. In thermal heat flux flowing to the earth's surface, there has also been an increase in measurements. In 1945, there were only 15 measurements; by 1965, the number had risen to 2,000, of which 90 percent were from the ocean bottom Another index of data flow is from the Upper Mantle project; 1,600 scientists from 50 countries are participating in this study in the 1963 to 1970 time period. Geochemical data flow is expected to increase for the following reasons: There is a growing research in: 1) trace elements, 2) solidification of molten matter into rocks, 3) organic geochemistry, 4) continental and oceanic heat flow, 5) magnetic reversal patterns, 6) new minerals exploration and techniques for their identification and recovery, and finally, 7) isotopic dating.

3. Data Flow

In describing environmental and geoscience data flow, the same breakdown of data is used. First, meteorological and related data are described; then data flow is described for hydrological, and finally, for geoscience data.

a. Meteorological Data Flow. The principal classes of data users and the use requirements are as follows:

- The Military - Data are required to support operations, research development, and test for the following general areas:

  (1) Warning and intelligence - atmospheric refraction and density for satellite tracking and ballistic missile warning; ionospheric variations for over-the-horizon propagation in tropscatter communications and backscatter and forward scatter radar; space environment effects on satellites;

(2) Weapons development and testing - wind vector, stratospheric winds for fallout dispersion and high altitude aircraft flights; sea state for ASW and SLBM testing, atmospheric density for nuclear effects analysis; low level meteorological conditions for tactical weapons testing; albedo, night glow and moonshine for night vision technology; reentry vehicle testing, chemical-biological warfare testing, launch vehicle takeoff stability, acoustic noise propagation, aircraft turbulence stress analysis; climatology for equipment environmental qualification testing;

(3) Tactical and strategic operations - cloud cover and height, winds, climatology;

(4) Logistics - weather forecasting and climatology.

- Civil Aviation - Data are required to design commercial aircraft and support operations. These include Clear Air Turbulence, low level turbulence, fog dispersal for all-weather landing, jet stream flow, all forms of weather forecasting. Other transportation, such as trucking and shipping, require weather forecasting and climatology data.

- Commerce - weather forecasting and climatology for operations and facility location planning and design.

- Agriculture and Natural Resources - weather forecasting, hurricane, . storm and flood warning, climatology data.

- Industry - weather forecasting, climatology, storm and flood warning, environmental aspects of capital equipment performance and product line design; air pollution.

- Maritime Transport Community - weather forecasting and surveillance, climatology.

- Scientific Community - climatology, environmental behavior, analysis and prediction, air/sea and air/land interfaces, upper atmosphere composition and behavior, solar emissions, tropospheric wind, temperature and humidity studies, planetary and tropical circulation, data correlations of the International Geophysical Year, the Global Atmospheric Research Program.

- Municipal and State Governments - weather forecasting, climatology, air pollution surveillance.

The principal sources and generators of meteorological data are as follows:

- Weather Forecasting and Surveillance (The World Weather Watch) - On a global basis, from nations supporting the World Meteorological Organization. Three WMO centers exist: Washington, D.C., Melbourne, Australia, and Moscow. From the military, there are the Naval Environmental Data Network and the Air Force's Digital Automated Weather Network, both of which feed the U.S. Weather Bureau's facility at Suitland, Maryland (the Suitland facility wears three hats: it is the North American regional center, and it also receives satellite data pertaining to weather). Satellite research applications to weather forecasting are handled via NASA-Goddard and the Air Force Cambridge Research Labs, as well as Naval Research Labs. (See Figure II-I-1.)

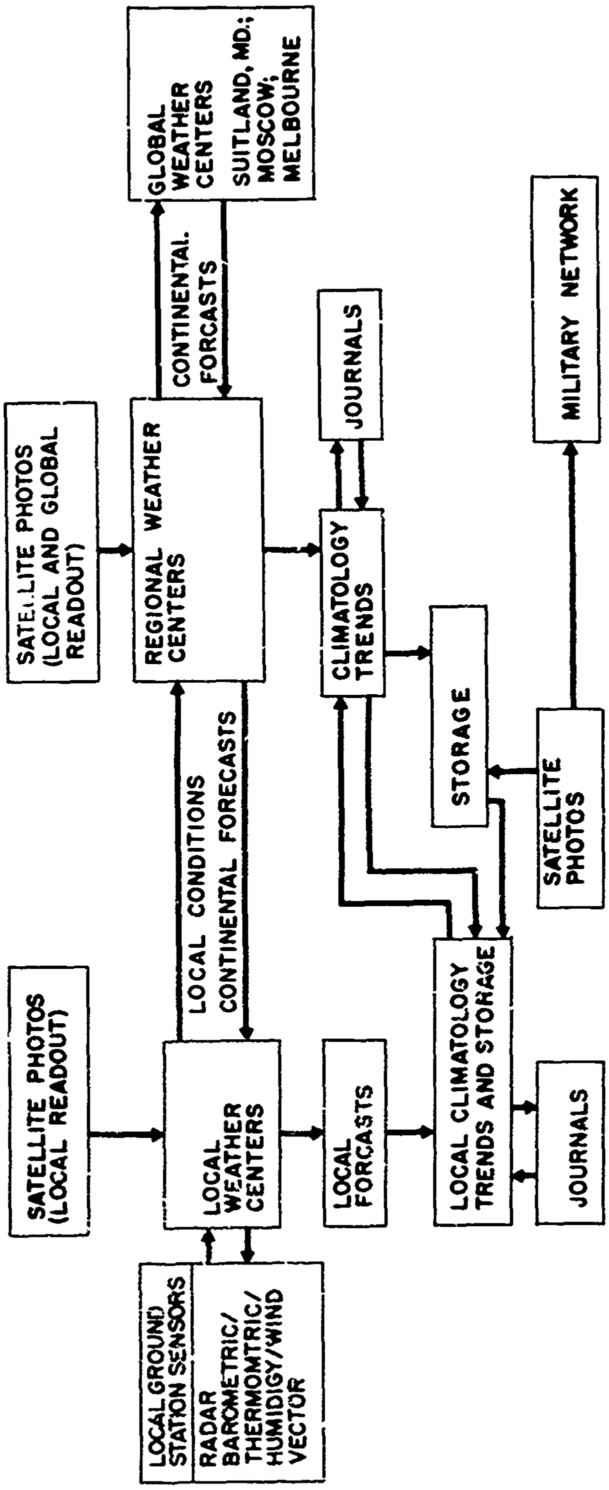


FIGURE II-I-1. WEATHER SURVEILLANCE AND FORECASTING DATA FLOW PATTERN

- Solar Flare Forecasting and Research - Centers include the National Center for Atmospheric Research at Boulder, Colorado; the Air Force Solar Flare facility at Ent AFB in Colorado Springs; and a naval facility at Corona, California. Observations in Switzerland also contribute solar flare data, along with Kitt Peak in New Mexico. Considerable solar particle research is handled at NASA-Ames at Moffet Field, California, as well as various universities.

- Aeronomy - National Center for Atmospheric Research, NASA-Goddard, Southwest Institute for Advanced Studies, Air Force Cambridge Research Laboratories, International Geophysical Union, various universities under National Science Foundation funding, U.S. Army Signal Corps, Smithsonian Astrophysical Observatory, ESSA-Institute for Telecommunications and Sciences and Aeronomy, National Bureau of Standards, aurora studies from the University of Alaska.

- Air Pollution - ESSA, HEW, various industry and university laboratories.

The principal intermediaries and data storage efforts for meteorological data are as follows:

- Weather Forecasting and Surveillance - Data are collected from over 11,000 ground stations and are fed to the National Meteorological Center at Suitland, Maryland. The raw data are supplied by networks supported by the Department of Commerce (ESSA) and the military networks, such as the Air Weather Service and the Naval Environmental Network. Other overseas inputs come from the World Weather Watch participants,

as well as direct lines to selected areas in the tropics (Carribean) and the far Pacific regions which have an effect on U.S. west coast weather. The raw data, in numerical form, are then fed to computerized models, which then automatically draw a nephanalysis for the North American Continent. The nephanalysis is then reissued to the majority of participating networks and sensor stations, which then make differential compensations for local conditions. These data are highly perishable, but for climatology purposes, they are stored at the National Weather Records Center in Ashville, North Carolina. In addition to the continental numerical forecasts, satellite cloud cover photos are fed to local areas directly from the ESSA satellites. Local radarscope records are also maintained for selected meteorological phenomena.

- Planetary and Tropical Circulation Research Data - These data are essentially the elements of the World Weather Watch plus the Global Atmospheric Research Program. GARP consists of the Line Island Experiments, which are nearing completion, and planned Barbados sea/air interaction exercises. In addition, the output of the ESSA, ATS, and Nimbus satellites are a continuing source of planetary circulation research data. These data are being analyzed by Air Force Cambridge Research Labs, the University of Chicago, the University of Wisconsin, ESSA's National Environmental Satellite Center, NASA-Goddard. GARP activities are coordinated by ESSA, the World Meteorological Organization, and COSPAR. Data storage is at the World Data Center-A on Meteorology (which is also the Nimbus Data Records Center and the National Weather Records Center) in Ashville, North Carolina and the World Data Center-A on Oceanography at Washington, D.C. (See Figure II-I-2.)

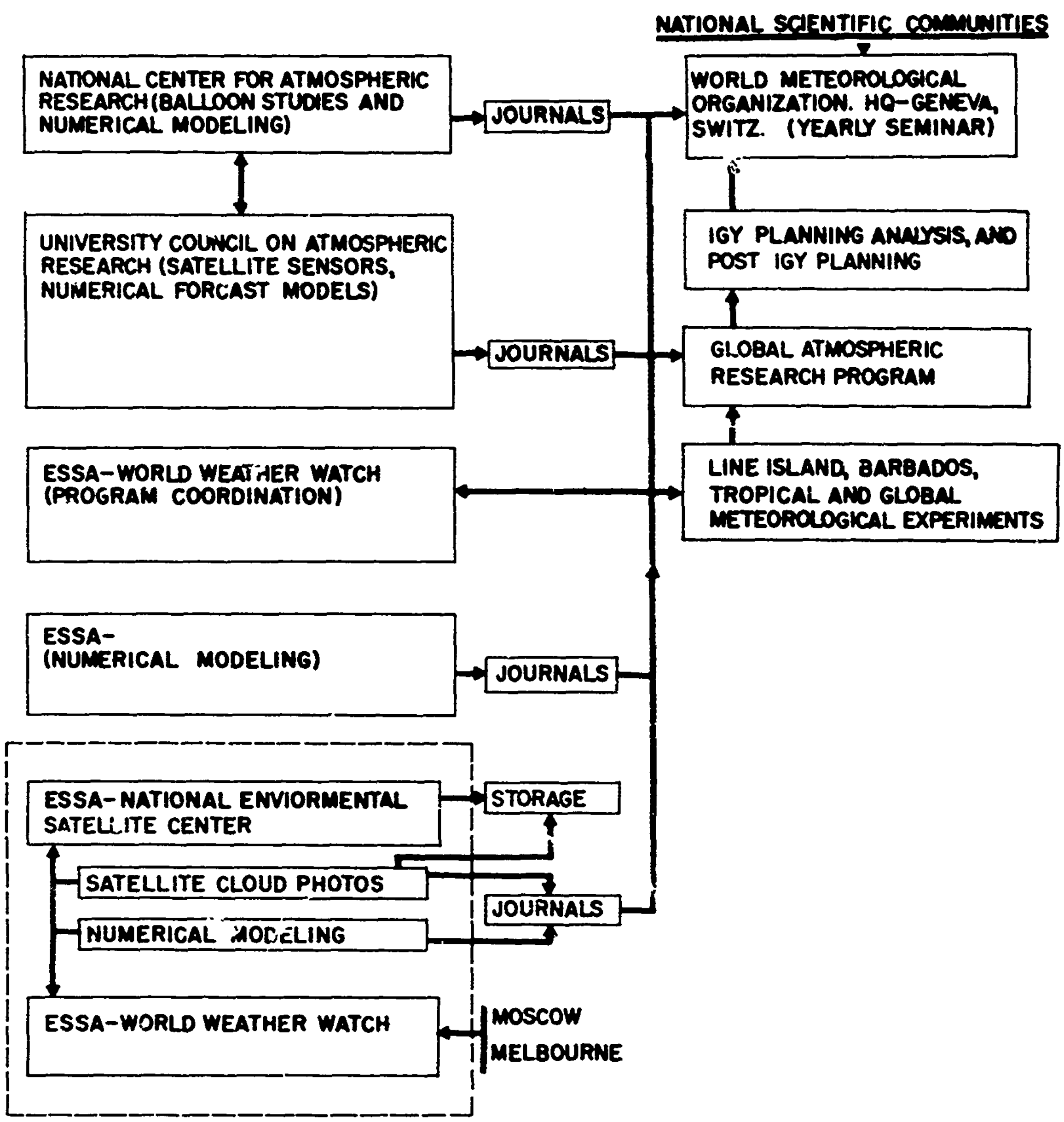


FIGURE II-I-2. PLANETARY CIRCULATION RESEARCH/OPERATIONS

- Aeronomy - Data collection includes a wide variety of sensors mounted aboard sounding rockets, balloon platforms, and satellites. Data are collected by the various telemetry networks (NASA's STADAN and the Air Force's Satellite Control Network of Sunnyvale) and analyzed by various governmental agencies. These include Air Force Cambridge Research Labs, NASA-Goddard Space Flight Center, NASA-Wallops Island, NASA-Ames, ESSA's National Center for Atmospheric Research, Southwest Research Institute, the Smithsonian Astrophysical Observatory, and the ONR/Naval Research Labs.

The question of data intermediaries and flow for airborne and spaceborne platforms is an extremely involved process. An index of this complexity is evident by the flow of telecommunications aeronomy data; it is more specialized, but the information that is available indicates the order of magnitude of the complexity of aeronomy data flow. For instance, ESSA's Institute for Telecommunications Sciences and Aeronomy data services indicate a wide range of activities. They include: four exchanges per day of solar-geophysical data from U.S. Observatories with 10 U.S. government agencies and five foreign data centers; three similar exchanges per day from overseas observatories with 40 U.S. Government agencies and 39 civilian scientists. In addition to the daily services, a number of monthly and quarterly bulletins are issued by ITSA. And for data banks, the World Data Center-A is fed by ITSA. Inputs include Ionospheric, Airglow, Solar Activity, Cosmic Ray, and Aurora data.

Generally speaking, the data flow for aeronomy data is one starting with the sensor, passing through the collection agency according to its function and program. The intermediaries, as shown in the case of the ITSA, represent government and civilian scientists who then contribute to the World Data Center-A collection with processed as well as raw data.

In the case of purely satellite traffic, NASA data are collected by the STADAN at Goddard, or Deep Space Network at Jet Propulsion Laboratory. The prime data receivers, however, are the Fairbanks, Alaska and Rosman, North Carolina tracking stations. Magnetic tapes are then transported to Goddard (some, however, are sent via data link), where they are stripped of their telemetry format via the STARS I and II data processors. The decommutated data are then submitted to additional computer processing in order to convert raw data into meaningful results.

b. Hydrological Data Flow. The complexity of data flow in this field is almost as great as that in meteorology. The flow pattern is summarized in Figure II-I-3. The principal classes of users are as follows:

- Steel and Metal Working: water pollution abatement, sedimentation of mill scale and solids ejected into water, electrokinetic methods for measuring concentrations of contaminants, water quantities available for cooling and flushing.

- Petroleum Refining: water use, water pollution by well drilling, water treatment, reuse, and industrial waste management.

- Power generation: water flow for cooling, generation of power, transportation of raw and waste materials, water storage technology, corrosion effects, economics of water resources within power distribution areas.

- Chemical industry: water analysis, purification, pollution, distillation, filtration, settling rates of contaminants, economic, legal aspects of pollution.

- Municipal: operation and management of public water works, sewerage disposal technology, water supply and distribution technology, leak detection techniques, trace contaminant technology, rainfall, consumption forecasting, public health.

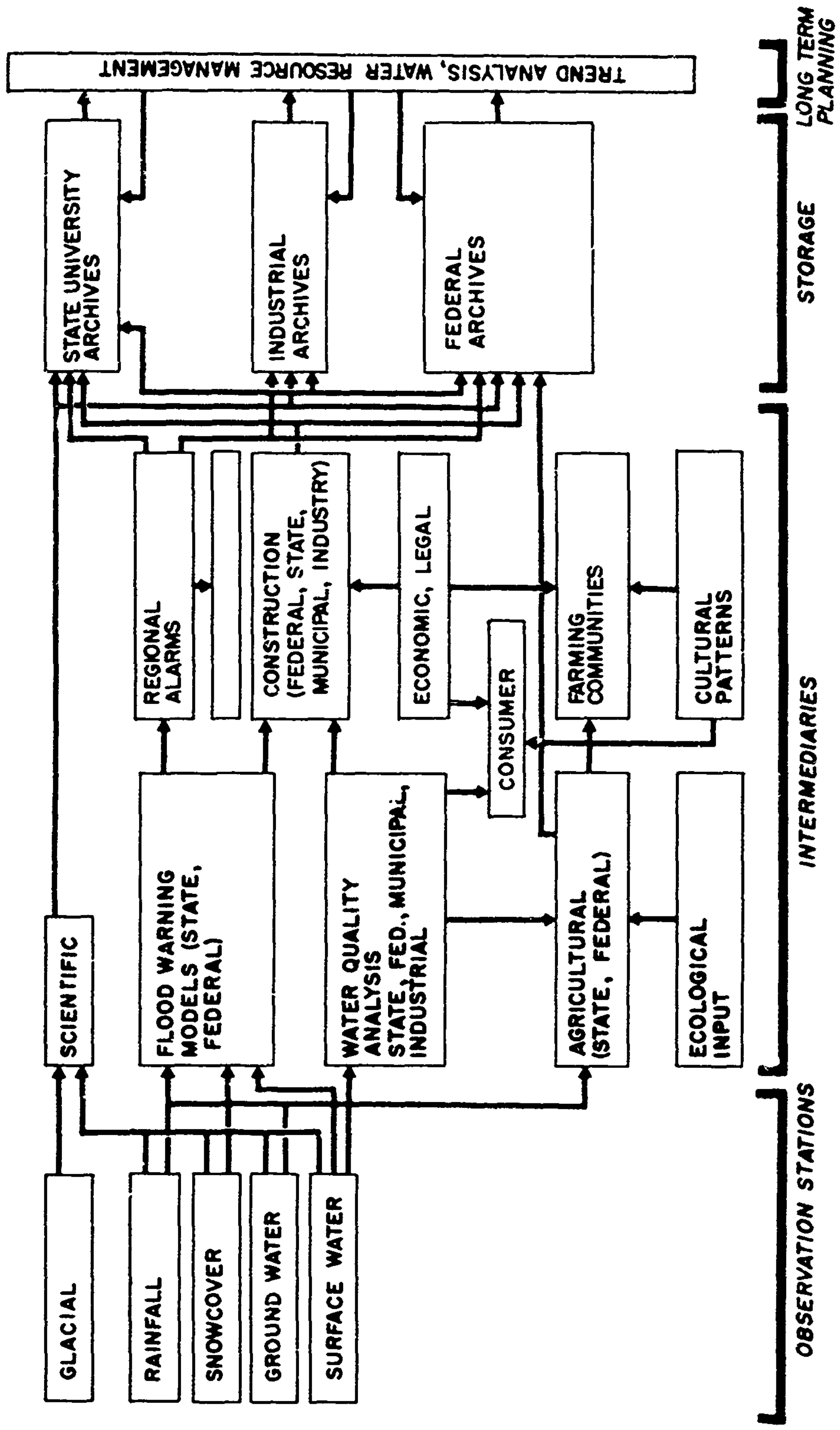


Figure II - I-3 Hydrological Data Flow

- State and regional: ground and surface water flow and quality, mine and industry acid runoff, salt water intrusion, hydrological flow modeling, flood warning, pollution control enforcement, irri gation, watershed conservation, rainfall, water basin development, public health.

- Federal:

  (1) Atomic Energy Commission - water desalinization with atomic energy, discharge of nuclear wastes, isotope for ground water traces, reactor cooling technology.

  (2) Department of Agriculture - watershed management, water to crop yield data, irri gation technology, soil leeching, erosion control, pesticide dispersal, climatic trends, crop resistance to drought and frost, water/chemical balances, drainage technology, farm waste removal, micro-biota ecology in agricultural waters.

  (3) Department of Commerce - highway drainage, census reports, trend data for water use and its relation to the economy, capital investments in water resources and management equipment.

  (4) Environmental Science Service Administration - metrology, hydrology sciences, mapping, climatology, instrumentation, hydrological modeling, trend data on extremes of rain and snowfall, river water level forecasting, all phases of hydrological cycle.

(5) Department of Defense - hydrologic surveys, sanitary engineering, water purification technology, water interaction with geologic surface features, construction, soil bearing strengths, micro-biota ecology, sewerage treatment, river and harbor engineering, mapping, removal of chemical and radiological agents from water supplies, advanced technology waste byproducts from exotic materials.

(6) Health, Education and Welfare - water ecology of communicable diseases, pollution assessment, fluoridation, aquatic life response to polluted environments, control of aquatic and semi-aquatic vectors of diseases, pesticide dilution, public drinking supplies, including technology and practice of distribution, multiple use watershed evaluation.

(7) Department of the Interior - water requirements of the mineral industry, water flooding of oil fields as a recovery technique, mine acid water drainage, hydraulic mining, recreational uses of water, reduction of flood hazards, runoff forecasting, stream oxygen depletion, reservoir seepage and evaporation loss control, limnology, eutrophication, river basin modeling, computer storage and retrieval of water quality data, aquatic life studies, hydraulics, hydrodynamics, water reclamation, fish kill statistics, solar heat distillation, membrane process technology, ion transfer, gaseous capacities, heat transfer.

(8) Non-Profit Laboratories - aquatic biology, water absorption, . hydration effects, solvent equilibriums, gamma ray spectrometry, free radicals, analog computer modeling, high and low velocity flow, cavitation, water quality, heavy water behavior, reactor cooling, radiation shielding, saline high energy reostats, meteorological activities.

The principal sources of hydrological data are Federal and state governments, as well as universities, non-profit laboratories, trade associations, and industries. This represents about 650 to 700 agencies involved in hydrological data, not including municipal and county governments. On a state level, available figures indicate between 200 and 1,000 ground and surface water stations are the primary measurement points, while on the Federal level, figures available indicate 7 to 10 thousand surface stations and over 500,000 ground stations feed the data networks. From these figures alone, one might conclude that the Federal effort is equal to the sum total of the state and regional effort. More detailed data are available in A Directory for Information Resources in the United States (Water), available from the National Referral Center for Science and Technology. . Nevertheless, for a gross inspection of the relationship between water flow, types of data, and agency type, the matrix is revealing. Considerable overlap is indicated while, in other cases, clearcut distinctions are shown.

Any attempt to describe the intermediaries in the flow of hydrologic data must recognize the existence of several means of describing the flow; the question of who is the intermediary then becomes a philosophical point. If the data are fed to a hydrological model (maintained by the state or Federal government), they have a flood forecasting role and the end user is the industrial and civil community affected by the flood warning. If these same data are used for reclamation and flood control construction, the intermediary is the Federal and state government, and the archives become the end of the data cycle.

If the sensor data are convoluted by agricultural thought processes, the Federal and state governments are intermediaries, but the agricultural community is the end user. This question of the same type of data having many uses characterizes much of the hydrology effort. But there is a distinction between the same type of data and the same data; obviously, each group intermediary draws his data from a different geographical area and with a different sampling rate.

The dissimilar use of similar data holds for the flow type data; when the question of purity and chemical quality enters the picture, it becomes more difficult to describe the data in any way except that they deal with pollution and that, to adequately describe the chemical composition of water entering the sea, one would have to identify about 500,000 trace components.

Thus, when considering the two main aspects of hydrological data - flow and composition - it is possible to envision many feedback loops between the sensors and the ultimate libraries of data.

c. Geoscience Data Flow. In this area, the principal classes of data users are:

- The military community, which utilizes geodetic data for ballistic missile targeting, space defense, orbital parameter corrections on spacecraft, whether for military or scientific missions; magnetic and gravitational anomaly data for anti-submarine warfare; cartography for reconnaisance, harbor maintenance and military base installations; cartography for tactical warfare; geologic data for base installations and tactical warfare; seismic data are needed for nuclear explosion detection and in tactical warfare, intruder penetration;

- The oil and mining community utilizes geophysical data for exploration; the data are generated by magnetometers and seismometers from airborne, tracked vehicle and shipboard stations; satellite remote sensing of gross surface features will be used in the future. Geochemical data are becoming more important for both fuel and mineral prospecting;

- The scientific community requires all forms of "Geo" data to study aging, motion, evolution of the earth, as well as magnetic and gravitational field data for space science and application;

- The aerospace community requires gravitational and magnetic field data to design propulsion and station-keeping equipment for manned and unmanned satellites; and

- Local, state, and Federal government require cartographic data for urban development, industrial planning, and watershed development.

The principal generators of geoscience data include the following:

- Federal Sources and their Contractors.

  (1) Cartographic and geographic data - Department of the Interior, U.S. Geological Survey; Department of Commerce, Coast and Geodetical Survey; Department of Defense, U.S. Army Mapping Service; Department of Transportation, Department of Highways;

  (2) Seismic Data - Air Force Cambridge Research Laboratories; Project VELA, University of Michigan; U.S. Coast and Geodetic Survey; U.S. Geological Survey; Atomic Energy Commission; Department of Housing and Urban Development;

  (3) Magnetic Data - U.S. Coast and Geodetic Survey; U.S. Geological Survey; U.S. Navy;

  (4) Geochemical Data - U.S. Bureau of Mines, Atomic Energy Commission, Private Sources;

- Cartographic and Geographic Data - Local chambers of commerce and state industrial development commissions, transportation industries; mining and petroleum exploration industries;

- Magnetic and Seismic Data - Mineral and petroleum exploration industries, universities; and

- Geochemical Data - Mining and petroleum exploration industries, industrial waste disposal oriented efforts, universities and non-profit research firms.

Science Communication
Washington, D. C. 200 07
COSATI Data Activities Study
Final Report - F44620-67-C-0022 30 April 1968

The flow of geoscience data varies with application; a sampling of seismic data flow cases illustrates this point. Commercial petroleum exploration entails gathering seismic (and magnetic) data at field locations on magnetic tape; the tapes are transported to the individual companies' data processing facilities. There are cases where the raw tapes from several firms are processed at a central location; these are for special services provided by Texas Instruments and IBM, etc.; elaborate security schemes insure that employees cannot identify the customer, and in certain cases, the location of the seismic profiles that are being processes. Just where proprietary considerations and the need to cooperate between firms cross the line is not evident from the literature as evidenced by a new data center in Dallas, Texas. This center, known as the Earth Science Data Center, is expected to employ 2,000 personnel to handle, store, and retrieve magnetic tape data and well drilling core samples. The Dallas center is largely a result of the efforts of the American Association of Petroleum Geologists and the Dallas Chamber of Commerce.

The contrast in seismic data flow for military applications is illustrated by the VELA network and a seismic signature analyzer developed by Air Force Cambridge Research Labs. In the VELA network, with its 525 element Large Aperture Seismic Array and several smaller conventional facilities operated by universities, all data are restricted to the military and the associated scientific community, in spite of the fact that it is the most advanced seismic system of its kind. The University of Michigan provides the VELA data center function. The micro-seismic noise detection system, on the other hand, is a small portable field instrument whose data are eventually fed to a computer. The seismic noise signature data are extracted from areas where motion-sensitive equipment such as phased array antennae and inertial guidance alignment platforms are located; both of these applications are classified, and the micro-seismic noise data are restricted. However, some applications, such as crystal growing, may be unclassified and make the data available.

TABLE II-I-6

ENVIRONMENTAL DATA SERVICE (ESSA)

GEOMAGNETIC DATA FILES	1. A microfilm file of geomagnetic data since 1957 from domestic and foreign geomagnetic observatories (approx. 250). Included in this file are: a. Copies of magnetograms from the observatories; b. Hourly value tabulations of the magnetic components at the observatories; c. Magnetic activity indices measured at the observatories; d. Tellurigrams and telluric hourly values from approximately 25 observatories; e. Listings of special events from approximately 40 observatories. 2. A file of magnetograms from observatories operated by the U.S. for years prior to 1957. 3. A magnetic tape file of hourly values of the magnetic components at all U.S. observatories since 1952 and for selected years prior to 1952. 4. A magnetic tape file of 2.5-minute values digitized from magnetograms since 1964 from a global distribution of approximately 60 geomagnetic observatories. Data are also available for selected observatories and intervals for 1961-1963. 5. A magnetic tape file of land, airborne, and marine absolute geomagnetic observations. This file consists primarily of component measurements (declination, horizontal intensity, vertical intensity, etc.) for points throughout the world. 6. A magnetic tape file of hourly values of equatorial D_{st} (storm-time variation) for most years since 1957.
SEISMOLOGICAL DATA FILES	1. A microfilm file of the daily seismograms since 1962 from a world-wide network of approximately 125 seismograph stations. 2. A file of seismograms from stations operated by the United States for years prior to 1962. 3. A magnetic tape file of earthquake epicenters since 1950. The file contains such information as time, location, magnitude, intensity, and damage. An incomplete file exists for years prior to 1950. 4. A magnetic tape file of a global distribution of P and S wave arrival times for earthquakes since 1961.
HYDROGRAPHIC DATA FILES	1. Fathograms, sounding tabulations, descriptive material, navigational aids, etc., for coastal regions of the U.S. 2. A punched card file containing all the essential information for approximately 40 C&GS nautical charts. This file will eventually include digitized data for all the C&GS nautical charts.
GEODETIC DATA FILE	1. A file, in printed form, of descriptions of all horizontal control points observed by the Coast and Geodetic Survey. 2. A file, in printed form, of descriptions of all vertical control points observed by the Coast and Geodetic Survey. 3. A magnetic tape file of all horizontal control points in the United States from all sources. This file contains such information as geographic position, name of control point, and source of data.

Earthquake warning activities are centered in the California-Alaska-Hawaii region. The National Center for Earthquake Research at Menlo Park maintains 30 seismographs; six stations are maintained by the Earthquake Mechanism Laboratory, and at Cal Tech, Earthquake Research Affiliates provide a commercial service for subscribers (utilities and the like that have dispersed unattended facilities). On a global basis, the Coast and Geodetic Survey monitors seismographs from 61 countries. Global, as well as regional, seismic events and their bearing on the inner structure of the earth are reported in the scientific journals. World Data Center-A also serves as a repository for seismic data. Seismic data on file at the Environmental Data Service appear in Table II-I-6.

The most economically significant use of geoscience data is for oil and mineral prospecting. Fundamentally, the search for oil and mineral deposits involves the same techniques and types of data. The first step is a preliminary search for and evaluation of available data contained in maps and publications. The purpose of this search is to discover geological, geophysical, and other anomalies that would indicate unusually high concentrations of chemical or mineral species of economic value. The search progresses from examination of coarse data concerning large regions to smaller favorable regions and finally, to promising prospects. The first data bases used are the map and publication resources of the U.S. Geological Survey (USGS) and the proprietary resources of private firms. The massive volume of geochemical and geophysical data generated by the Experimental Geology program is not presently archived in a data system; moreover, there is no plan to establish such a data base because of the enormous cost which should be borne by the primary benefactors, the earth resource industry, rather than the taxpayer.

The data contained in this national resource (maps and publications) include both directly and indirectly obtained data. Indirect data include geophysical surveys (magnetic, electromagnetic, radioactive, gravity, seismic, and thermal gradient data both from the air and at the earth's surface) and geochemical and geobotanical surveys. Direct data include geologic and photogeologic maps, ore guides (data on enriched areas and patterns thereof), data on panning, trenching, pitting, drilling, or geochemical sampling. USGS makes these data available on request on an equal basis in response to all queries. State geological services sometimes go a step further and provide interpretation of data which can lead a prospector or prospecting organization to specific promising regions.

Where initial searches of available data resources provide hopeful evidence of minerals or oil, further data development is warranted. Both the direct and indirect types of data are gathered for the most promising of the smaller regions. For mineral and /or metallic ore prospecting, data other than geophysical data are used because of the relatively small size of most ore bodies and because of the great variety of their size, physical properties and the influence of their highly variable geological environment.

The mineral geosciences, due to their historical precedents and economic value, contribute heavily to the geological arts; their data are best estimated by the number of worldwide organizations and journals. This figures on 137 countries, 350 worldwide organizations, and over 400 worldwide journals. The total number of soil and mineral samples stored within the auspices of these organizations is not available from easily accessible sources, but it must be extremely large. A figure is available on ocean bottom cores; the Lamont Geological Observatory has collected over 4,000 cores (probably varying from 50 to 70 feet in length) from 45 expeditions.

For geomagnetic surface data, the U.S. Coast and Geological Survey monitoring effort provides a figure on the scope of these data. Over 150,000 stations have been measured on a worldwide basis. Over 300 worldwide locations contribute to World Data Center-A, maintained by ESSA's Coast and Geodetic Survey at Rockville, Maryland. The readings vary in sampling frequency, but many are on an hourly basis throughout the year. The data are stored in various forms; microfilm, magnetic tape, publications, and bulletins. Details of geomagnetic data on file at the Environmental DataService appear in Table II-I-6.

The most prolific contributor to geodetic data is from space-oriented activities. Datum plane linkage runs about 100,000 sightings per year with 10 to 15% usable data yield. These data only require triangulation and other algebraic and statistical manipulation to measure datum plane linkage; about 185 worldwide sights are participating, and the results are available from professional society journals. On the other hand, determining the shape of the geoid and determining the spherical harmonic coefficients requires much more effort. All satellites and all combinations of orbits are of value;

consequently, there is varying emphasis on low and high mass/volumeratio sightings (to separate drag components from gravitational perturbances), as well as polar vs. equatorial orbits. Several agencies are cooperating under the auspices of the National Geodetic Satellite Program. These agencies include the Air Defense Command with its 496L Spacetrack net; the Smithsonian Astrophysical Observatory, with its 12 camera stations, the Navy's TRANET, NASA's STADAN net, and also cameras operated by the Coast and Geodetic Survey (61 stations were used in sighting). Sighting loads are factors very high; the 496L network makes 400,000 sightings per month of 1,250 space objects, all of which require geodetic data for orbit prediction. A small percentage of these sightings are used for geoid determination, although all must be correlated against drag and geoid terms. One satellite with an optical beacon (GEOS-A) provided over 15,000 sightings in a one-year period. The results of geodetic satellite sightings are reported in the scientific literature. NASA-Goddard, which manages the NASA training networks, the SAO, and the Air Force at Colorado Springs are repositories for geodetic data; both processed and raw. The Navy maintains a computation and analysis center at Dahlgren, Virginia. Details on the Environmental Data Service data files on geodesy appear in Table II-I-6.

Cartography efforts are led by the Army Map Service, the U.S. Coast and Geodetic Survey, and the Department of the Interior. These efforts serve as primary data bases for national, global, and regional uses. The Bureau of Reclamation, for instance, distributed 48,000 copies of bulletins and publications in 1959. Local data, on the other hand, are often generated by state and local governments, as well as commercial firms.

4. Principal Issues

The main point that requires examination in environmental data management is that "data are data," and become information only when interpreted by the various disciplines. There is adequate reason to believe that some data straddle disciplines; if this is so, sensor resolutions, frequency of measurement, and temporal, spectral, and spatial domains should be indexed in abstracts. This would allow interdisciplinary access to data, and would relieve the cries of "inability to retrieve data" and prevent generation of additional data by those investigators who are in the best position to resolve the problem.

Calibration data and standardization should also be emphasized. Much of the future satellite data credence will require correlation with ground truth sites. Ground truth data are now being gathered by aircraft; and they serve as a baseline for future progress and should be included in any future remote sensing data banks. Similar arguments can be found for other types of environmental data.

The large amounts of data - in all forms - indicate that extremely informal networks must exist for filtering and sorting the data. Users must maintain professional relationships via societies, trade fairs, and the like. Data processing, an intermediate but expensive step between sensors and tabular/graphical results, is strangely absent in the literature. Many software programs, if grossly identified, could be applicable to expanded usage if proprietary or security considerations do not intervene.

The last point of general consideration in environmental data flow analysis is the interplay of the quest and the instrumentation tool. The instrumentation determines the quality of the data, while the quest determines its volume via spatial and temporal domain requirements. Thus, the importance of the quest cannot be underrated when studying the flow and utilization of environmental data.

Examination of the more specific fields of aeronomy and meteorology has shown that rather large amounts of data are presently being generated, but yet no clear picture is available from the literature as to the true scope and magnitude of the present effort. This is not implying that such information cannot be collected and analyzed; indeed, it can, as one continues to pursue "tracer" facts, facts that allow conversion and extrapolation to the desired results of such a study. Notwithstanding present inability to define the present scope of meteorological and aeronomy data (an inability due mainly to time and establishment of analytical techniques to achieve the data survey goals), future programs would seem to inundate to levels we presently cannot imagine.

A point of clarification is badly needed on the utilization of data from meteorological satellites such as Nimbus, ESSA, and ATS. These data are piling up; in some cases, they are duplicated under the guise of operational versus research data. The fact of the matter is

that most operational data that are stored are really research data, even though ESSA is an "operational" bird, while Nimbus is a "research" bird. The ESSA data are more continuous than the Nimbus data, but the Nimbus data are of much better quality. Needless to say, a utilization survey should be run on meteorological satellite cloud photos.

In retrospect, one must recall that there are four billion cubic miles of atmosphere to watch; for these reasons, improved data management is required if we are to move from the present scratching of the meteorological surface into meaningful, economically viable programs such as GARP and solar flare forecasting.

In the field of hydrology, so many agencies are involved in water resources, and their data holdings are so voluminous, that no standard format exists in data collection and storage. If standards were available, some of the questions of water flow and their interrelationship with biological and social effects could be attempted. Standardization of format would also permit synoptic studies, provide some possibility for synchronization of sampling.

A more likely advance in the standardization of hydrological data will come by way of the satellite. That is, satellite high resolution visual and infrared maps could provide the synchronizing factor to much of the flow data. Hence, the problem to be faced in the future, assuming the satellite program evolves as anticipated, is the distribution of its data to users, as well as their ability to utilize such data. Satellites have spotted ground water leakage to the sea; they have detected patterns indicative of sedimentation and pollution and they are attempting to provide flood level reports which may lead to advances in flood forecasting. Therefore, while it is difficult to pinpoint present data problems, the proper usage of the applications satellite should force some data management issues to surface in the future.

The principal problem in geoscience data management seems to be the question of what is available today and the realization that tomorrow will bring more data. A report on geodetic satellite data formats notes the need to periodically disseminate information to qualified users relative to the current inventory of observations, nominal orbit elements, information pertaining to the general location of all tracking stations, and plans relative to their movement. The message here is obvious: the sensors move continuously, and there seems to be no regularity to the dissemination of data.

Geophysical data for prospecting are also worth singling out as a major problem. These data have been of increasing importance in prospecting for minerals because of the shift in emphasis from surface to submerged deposits. The development of airborne, and more recently, space-borne, measurements such as magnetic, electromagnetic, and radioactivity has made possible the accumulation of massive data coverage of broad regional areas at relatively low cost. The ultimate implementation of the National Space Science Data Center and its program for storage and dissemination of satellite-acquired geological and geophysical data will promote the trend toward increased use of geophysical data for broad-region mineral prospecting.

Thus, one might conclude that the principal problems are periodicity of distribution, format of distribution, which has been set by historical precedent, coordination of major measurement endeavors, and an inability to cope with the environmental data situation as it exists.

Science Communication
Washington, D. C. 200 07
COSATI Data Activities Study
Final Report - F44620-67-C-0022 30 April 1968

Selected Bibliography

1. "Learning to Live with Earthquakes". Business Week, December 23, 1967, p. 56.

2. Program of the 48th Annual Meeting of the American Geophysical Union, April 1967.

3. 1967 Tutorial Lecture Series, Institute of Environmental Sciences.

4. Responsibilities of the Technical Community and the Government in the Transfer of Information. Report by President's Science Advisory Committee, January 1963.

5. Study of Federal and Non-Federal Science Information Processing and Retrieval Program. Senate Document #113, U.S. Government Printing Office, 1961.

6. Space Observations and Earth Resource Studies. W.A. Fisher, U.S. Geological Survey.

7. "Global Weather Forecasting". Space/Aeronautics, October 1967.

8. Proceedings of the World Meteorological Organization Conference on the Global Atmospheric Research Program, Stockholm, July 1967.

9. Scientific Information Activities of Federal Agencies, National Science Foundation, Publication 61-77.

10. "Earth Resource Surveys", Space/Aeronautics, July 1958.

11. Mining, Minerals, and Geosciences, S.R. Kaplan, Interscience/Wiley.

12. Observational Data Formats and Identification of Supplementary Documentation for the National Geodetic Satellite Program, Kahler/Roy - Wolf Rand Corporation.

13. Land Use Information, G.R. Hall - Rand Corporation, Clearinghouse Accession Number AD 642 282.

J. Oceanography

1. Introduction

For the purpose of this study, oceanography is defined as the investigation of the oceans, other salt water bodies, and fresh water bodies including estuaries. It includes: the study of the internal and external forces that cause water motion; the interaction of water bodies with each other and the atmosphere; the taxonomy, ecology, and dynamics of the marine biosphere; the chemistry of the waters and water-body bottoms; and the geophysics and the physics of the marine and interfacing environments; as well as the relationships between these several factors. Obviously, it is impossible to separate the field of oceanography from the geosciences (i.e., geophysics, geochemistry, geology); so those aspects of geoscience which are relevant to oceanography are treated in this study, and those relevant to dry land are treated in the foregoing section on "Environmental and Geosciences." Those data which are relevant to both water-cover and dry land are considered geoscience data, and accordingly are discussed in the foregoing section.

In recent years, the basic science of oceanography has drawn the interest of the commercial and governmental organizations which utilize or exploit the oceans' resources. These users of the ocean resources have become increasingly sophisticated and are, in fact, utilizing data which parallel the quality of that used by the oceanographic phenomenologist. In studying national data activities, we have observed that practical application of data developed by and for the phenomenologist is increasingly used in many technological communities (e.g., the food and paper industries) as their arts and crafts methodologies are maturing to become more scientific. The impact of this trend is an overwhelming increase in demand, generation, and use of phenomenological data in an operational context, and an increasing need for organized management of data of interest to an increasing number of scientific and technological communities. This increasing demand for oceanographic data results from a recognition of its utility by the following nine communities of interest:

1. The oceanographic research and development community, which seeks advancement of knowledge concerning oceanographic phenomenology.

2. The defense and aerospace community, which utilizes air-sea interface data. undersea data, and ocean floor data to attain national strategic and tactical objectives;

3. The ocean engineering community, which is concerned with construction of stationary facilities at sea or over large bodies of water;

4. The health and welfare communities, which are concerned with the effects of estuarine and sea conditions on water and food resources;

5. The maritime and air transport community, which is concerned with the sea state and air-sea interaction aspects of meteorology as they affect the operations of this community;

6. The conservation community, which is concerned with management of the resources of the sea (such as water, wildlife, and plant life) insofar as they are of recreational and other value to the nation;

7. The commercial food and fisheries community, which is concerned with optimal exploitation of fish and other food resources associated with the oceans;

8. The mining and petroleum engineering communities, which are concerned with optimal exploitation of the mineral and fuel resources of the sea, its floor and subfloor; and

9. The marine engineering community, which is concerned with the design and operation of surface and submersible vehicles.

The importance of these objectives of the marine science community suggest that the research conducted in order to exploit the marine environment would constitute a large part of our national technical activities, particularly since the oceans cover more than 71% of our earth's total surface. However, despite its broad technical scope and obvious importance, the identifiable oceanographic effort constitutes only about 3% of the Federal scientific and technical program. Actually, the total budget presently projected for the 10-year period through 1972 totals about $2. 3 billion, including all hydrographic data activities in support of Naval Fleet operations. The estimated expenditures for the FY 1968 Federal oceanographic program are about $448 million and $516 million for FY 1969, which is about 10% of the space exploration program. Two years ago, the expenditures projected by the Interagency Committee on Oceanography for the period from 1963 through 1972 totaled up to about $2. 3 billion, including all oceanographic activities in support of the U.S. Navy Fleet operations. Over 50% of all Federal funding for marine sciences is budgeted by the Department of Defense and more than 80% of this support is connected with operations such as fleet support and antisubmarine warfare. Table II-J-1 summarizes the Federal support of oceanographic research and engineering development.

In evaluating the relative size and importance of the national oceanographic effort, it is important to consider the very large sector of activity implemented by petroleum exploration and exploitation firms, which do not disclose the magnitude of their oceanographic programs. It is conceivable that these firms might be operating oceanographic programs comparable in total size to the total Federal program, but it is also important to place these programs in proper context, inasmuch as the research findings are very largely proprietary and therefore may not properly be considered a nationally available resource.

Another industrial sector to be considered in evaluating the size and significance of the national oceanographic effort and the associated data management problem is the degree to which U.S. industry is involved in:

Table II-J-1

Federal Support of Marine Science and Engineering Development for FY 1969

National Objective	Federal Agency		Budget (millions of dollars)
International Cooperation	State Department		8.2
National Security	Department of Defense		150.1
Fishery Development and Sea Food Technology	Department of the Interior		42.7
Transportation	Department of Commerce	7.6	
	Department of Defense	3.0	
	Dept. of Transportation	4.8	15.4
Coastal Zone Development and Conservation; Shore Stabilization and Protection	Department of Defense		1.7
Marine Pollution Management	Department of Defense	2.4	
	Department of the Interior	6.3	8.7
Recreation and Conservation	Department of Defense	1.7	
	Department of the Interior	16.2	
	Dept. of Transportation	0.3	28.2
Health	Department of H.E.W.		6.0
Non-Living Resources	Department of the Interior		9.8
Oceanographic Research	Department of Defense	38.0	
	Department of Commerce	4.2	
	National Science Found.	36.0	
	Dept. of Transportation	15.7	
	Smithsonian Institution	1.4	
	Atomic Energy Comm.	4.4	99.7

(Continued)

Table II-J-1
(Continued)

National Objective	Federal Agency	Budget (millions of dollars)	
Education	Department of Defense	1.3	
	Department of Commerce	0.1	
	Department of the Interior	0.2	
	National Science Found.	4.7	
	Department of H. E. W.	1.5	
	Dept. of Transportation	0.1	7.9
Environmental Observation, Prediction, and Services	Department of Defense	11.3	
	Department of Commerce	6.3	
	Atomic Energy Comm.	0.8	
	Dept. of Transportation	6.8	
	NASA	1.3	26.5
Ocean Exploration, Mapping, Charting, and Geodesy	Department of Defense	72.3	
	Department of Commerce	19.5	
	NASA	0.3	92.1
General Purpose Ocean Engineering and Development	Department of Defense	14.9	
	Atomic Energy Comm.	6.6	
	Dept. of Transportation	5.3	26.8
National Data Centers	National Oceanographic Data Center	1.8	
	Smithsonian Oceanographic Sorting Center	0.3	
	Great Lakes Data Center	0.2	
	National Weather Records Center	0.1	2.4
		TOTAL	$516.2

- Research Vehicle Development, including design, development, manufacture and operation;

- Man in the Sea Research, including the design, development, manufacture, and use of scuba gear, diving systems, undersea test and habitation systems, and undersea manipulation gear;

- Major Equipment Components, including the manufacture of buoys, buoy systems, towing sleds, unmanned undersea platforms, recorders, instruments and samplers;

- Instrument Development, including the design, development, manufacture, and testing of samplers, sensors, recorders, and undersea TV;

- Communications Gear Design, including development of navigation, sonar, data transmission cables and connectors, and radio frequency equipment;

- Test and Analysis, including equipment calibration, environmental simulation and hydrodynamic studies;

- Survey and Research Services, including corrosion, geophysical, geological, geochemical, physical oceanography, and biological research;

- Structural Design, including design and manufacture of pressure vessels, deep submergence vehicles, buoyancy systems, and undersea platforms; and

- Construction, including undersea engineering, equipment search and salvage.

A survey of some 98 major domestic corporations published by International Science and Technology in April 1967 revealed that all of these firms were involved in at least one of these activities related to the field of oceanography, and that 26 of them were engaged in activities with budgets in excess of $1 million per year. Table II-J-2 summarizes the findings of the survey.

Table II-J-2

U. S. Industry Involvement in Oceanographic Activities

Manufacturing Firms	Fields of Interest											
	Major Systems	Hardware	Services (incl. R&D)	Research Vehicles	Man-in-the-Sea	Large Components	Instruments	Communications	Test and Analysis	Survey and Research	Structures	Heavy Construction
ACF Industries		x	x				x		x			
Allis-Chalmers		x	x	x		x						
Amer. Bosch Arma		x	x	x	x		x	x				
Barnes Engineering		x	x									
Bendix (Marine Advisors)*		x	x	x		x		x				
Borg-Warner		x	x	x								
Clevite Corp.	x	x	x	x		x		x				
Cohu Electronics		x	x									
Collins Radio	x	x	x	x		x		x				
Coors Porcelain		x	x			x					x	
Corning Glass		x	x	x		x					x	
Dynamics Corp. of America		x	x									
EG&G	x	x	x	x		x	x	x	x	x		
Edo Co~~r~~p.		x	x	x			x	x				
E.I. duPont de Nemours*		x	x	x				x			x	
Ford Motor (Philco)*		x	x	x			x	x				
Francis Associates		x	x				x	x				
General Instrument		x	x	x			x	x				
General Precision*		x	x	x		x		x		x		
B.F. Goodrich		x	x				x				x	
Hektor Scientific		x			x		x					
Hewlett-Packard		x	x	x			x	x				

(Continued)

*Annual expenditure over $1 million.

Table II-J-2 (continued)

Manufacturing Firms	Fields of Interest											
	Major Systems	Hardware	Services (incl. R &D)	Research Vehicles	Man-in-the-Sea	Large Components	Instruments	Communications	Test and Analysis	Survey and Research	Structures	Heavy Construction
Honeywell, Inc.*	x	x	x	x			x	x				
HRB Singer	x	x	x		x	x	x	x				
IBM	x	x	x	x				x				
Interstate Electronics	x	x	x			x	x	x	x	x		x
Joy Manufacturing		x						x	x		x	
Lear-Siegler		x	x	x		x	x	x				
3M Company		x	x			x					x	
Magnavox		x	x			x	x	x				
Mine Safety Appliance		x	x	x	x	x					x	
Moog, Inc.		x	x	x				x				
Motorola		x	x				x	x				
Olin Mathieson Chemical		x	x					x			x	
Penn Engineering & Mfg.		x	x			x	x					
Chas. Pfizer		x	x	x								
Pike Corp. of America		x	x						x			x
Radio Corp. of America	x	x	x				x	x				
Raytheon Co.	x	x	x	x			x	x	x	x		
Schlumberger Corp.*	x	x	x			x	x	x	x	x	x	
Sparton Corp.		x	x			x	x					
Straza Industries		x	x				x	x				
Texas Instruments	x	x	x	x		x	x	x		x	x	
Textron Corp.	x	x	x	x			x	x				
Titanium Metals Corp.		x	x								x	

(Continued)

*Annual expenditure over $1 million.

Table II-J-2 (continued)

Manufacturing Firms	Major Systems	Hardware	Services (incl. R&D)	Research Vehicles	Man-in-the-Sea	Large Components	Instruments	Communications	Test and Analysis	Survey and Research	Structures	Heavy Construction
U.S. Rubber Co.*	x	x	x		x						x	
Universal Match Co.		x	x	x			x	x				
Walter Kidde & Co.		x	x	x	x		x	x	x	x	x	
Western Electric Co.*	x	x	x	x								
Aerospace Firms												
Aerojet-General Corp.*	x	x	x	x		x	x	x	x		x	
Astropower, Inc.		x	x	x				x	x			
Atlantic Research		x	x		x		x	x				
Boeing Company	x	x	x	x						x		
Douglas Aircraft	x	x	x				x	x				
Grumman Aircraft Eng'g.*	x	x	x	x	x				x	x	x	
Hughes Aircraft	x	x	x		x	x	x	x				
Kaman Aircraft	x		x						x		x	x
Litton Industries*	x	x	x	x		x	x	x	x	x		
Lockheed Missiles & Space*	x	x	x	x	x	x		x	x	x	x	x
Martin Marietta Co.*	x	x	x									
North American Aviation*	x	x	x	x	x	x	x	x	x	x	x	x
Northrop Corp.*	x	x	x	x	x							x
Rohr Corp.		x	x						x			
Sperry-Rand Corp.	x	x	x	x								
United Aircraft Corp.*	x	x	x	x				x				

(Continued)

*Annual expenditure over $1 million.

Table II-J-2 (continued)

Heavy Industry	Major Systems	Hardware	Services (incl. R&D)	Research Vehicles	Man-in-the-Sea	Large Components	Instruments	Communications	Test and Analysis	Survey and Research	Structures	Heavy Construction
Aluminum Co. of America		x				x					x	
Bethlehem Steel Co.*		x		x							x	
Chicago Bridge & Iron	x	x	x		x				x		x	
Sun Shipbuilding*	x	x	x	x		x					x	
Mixed Industries												
Amer. Machine & Foundry	x	x	x		x	x	x	x	x			
Avco Corp.		x	x			x	x				x	
General Dynamics Corp.*	x	x	x	x	x	x	x	x	x	x	x	
General Electric Co.	x	x	x		x				x			
General Motors*	x	x	x	x		x		x	x	x	x	x
Reynolds Metals Co.*	x	x	x	x	x			x	x	x	x	
Vitro Corp. of America	x	x	x		x							
Westinghouse Electric*	x	x	x	x	x	x	x	x	x	x	x	
Oceanography Firms												
Alpine Geophysical Assoc.	x	x	x	x		x	x	x	x	x	x	
Benthos Co.		x	x	x		x	x		x		x	
Bisset-Berman	x	x	x	x		x	x	x	x	x	x	
Geodyne Corp.	x	x	x			x	x	x			x	
Hydronautics, Inc.*		x	x	x	x	x	x		x	x		
Marine Acoustical Serv.		x	x	x			x	x		x		x
Marine Advisors (Bendix)	x	x	x			x	x	x	x	x		

(Continued)

*Annual expenditure over $1 million.

Table II-J-2 (continued)

Oceanography Firms	Fields of Interest											
	Major Systems	Hardware	Services (incl. R&D)	Research Vehicles	Man-in-the-Sea	Large Components	Instruments	Communications	Test and Analysis	Survey and Research	Structures	Heavy Construction
Marine Technology Corp.		x	x	x						x		
Ocean Research Equip.		x	x	x		x	x	x	x		x	
Ocean Science & Eng'g.	x	x	x	x	x	x	x	x	x	x	x	x
Ocean Systems (UCC)*	x	x	x	x	x				x			
Perry Submarine	x	x	x	x					x			
Other Firms												
CBS Laboratories			x					x				
Global Marine Exploration	x		x	x								
Lunn Laminates, Inc.		x				x				x		
International Nickel*		x	x	x					x	x	x	
ITT*	x	x	x			x	x	x	x		x	x
Miller Highlife Brewing		x				x						
Utah Const'n. & Mining		x										x

*Annual expenditure over $1 million.

Source of Information: International Science & Technology, April 1967.

Science Communication
Washington, D. C. 200 07
COSATI Data Activities Study
Final Report - F44620-67-C-0022

30 April 1968

Realization of the importance of the national oceanographic objectives and increasing attention which is being directed at their attainment, was a primary motive in the legislation of the Marine Resources and Engineering Act of 1966. The broad set of action programs which this Act mandates are aimed at development, encouragement, and maintenance of a coordinated, comprehensive, and long-range national program in marine science for the benefit of mankind.

The program includes establishment of a Marine Sciences Council in the Executive Office of the President and an independent advisory Commission on Marine Science, Engineering and Resources to consist of 15 appointed representatives from government, industry and the private research institutes. In accordance with Public Law 89-454 and the changes enacted by Public Law 90-242, the Commission is to submit a report to the President by January, 1969, giving the joint recommendations of the private and government sectors concerning the management of the future national oceanographic program. In addition, the Council is to complete its policy and program implementation plans by June, 1969, and the recommended program should at that time be implemented.

A major consideration in the studies by the Council is information/data management. Accordingly, the Council has contracted with Systems Development Corp. to conduct a major study of marine science data management. This study is to be completed early in 1969. This study conducted by Science Communication, Inc., of data activities, provides a perspective of marine science data activity in a total science and technology context. The complete analysis of organizations, organizational involvement, data forms and formats, data processing techniques, data quantity, sensor standards, and nonscientific factors associated with marine science are the subject of the SDC study.

It is expected that the findings and recommendations of both the Council and the Commission will have far-reaching effects on the management of both Federally and privately sponsored marine science activities, including the data management and data system management aspects. Penetrating studies of national scientific and technical efforts, such as this specific probe focused at the marine science field, are the practical basis for formulation of specific data management policies and plans.

The plans that evolve out of the Council and Commission studies will take into account the changing interdisciplinary and inter-mission roles associated with the field of oceanography, particularly in data management and data systems planning. These roles are of significance in this study of national scientific and technical data efforts. Two principal trends should be noted with regard to the changing discipline relationships associated with the marine sciences. First, the spectrum of disciplines related to the marine sciences is broadening as the nine communities of interest, listed in the first part of this subsection on oceanography, are demanding wider varieties of technical effort and capabilities. These are illustrated in the following table.

The Spectrum of Disciplines Typically Related to Oceanography

Physical Sciences	Earth and Environmental Sciences	Life Sciences	Engineering
Acoustics	Geochemistry	Taxonomy	Undersea Acoustics
Optics	Geophysics	Physiology	Undersea Communication Systems
Fluid Mechanics	Geography	Biophysics	Submersible Structure Engineering
Electronics	Geodesy	Biochemistry	
Analytical Chemistry	Meteorology	Microbiology	
Organic Chemistry	Physical Oceanography	Ecology	
Physical Chemistry			
Inorganic Chemistry			

The second developing trend with regard to the marine science disciplines is the increasing recognition of physical oceanography as a discrete field of knowledge. This is illustrated by the increasing number of educational institutions which offer degrees in oceanography and the increasing number of oceanographers who hold degrees in physical oceanography rather than the related geosciences. According to a recent National Science Foundation survey, 12 U.S. universities now offer undergraduate and graduate programs in oceanography, and a total of 27 universities offer programs in the marine sciences. The following table lists the universities and the fields of study they offer.

University	Field
Texas A. & M.	Oceanography
University of Alaska	Oceanography
University of Chicago	Geophysical sciences
University of Connecticut	Marine sciences
Columbia University	Oceanography
Cornell University	Marine ecology
University of Delaware	Biological sciences
Florida State University	Marine sciences
Harvard University	Marine sciences
University of Hawaii	Oceanography
Humboldt State College	Fisheries
Johns Hopkins University	Oceanography
Mass. Institute of Technology	Oceanography
University of Miami	Fisheries, Marine biology, Oceanography
University of Michigan	Marine sciences
U.S. Naval Post Graduate School	Oceanography & meteorology
New York University	Oceanography
State University of New York Maritime College	Meteorology & oceanography
Oregon State University	Oceanography
University of Rhode Island	Oceanography
Scripps Institution of Oceanography	Marine biology, and oceanography
University of Southern California	Biology and geology
Stanford University	Marine biology
University of Texas	Marine science
Virginia Institute of Marine Science	Marine science
University of Washington	Oceanography
University of Wisconsin	Oceanography, and limnology

Prospects for the increase in number of curricula, institutions and academic enrollment in the marine science field are particularly likely with plans evolving for a multi-billion dollar International Oceanographic Decade to begin in the 1970's. This program as well as the plans and programs of the Marine Council and the Marine Commission portend enormously increased requirements for oceanographic and related data management activities.

2. Data Characteristics

There are five principal categories of data of concern in the field of oceanography. These are geology and geophysics, marine biology, marine chemistry, meteorology and climatology, and physical oceanography. As in other fields of science and technology the data characteristics are determined by the sensors or measurements used in data generation, and the mode in which they are used. Table II-J-3, which comprises the following six pages, lists typical classes of data used in each of the five fields, the primary measurements which generate the raw data, the artifacts which contain the data, some examples of the functional utility of data in the field, and typical derived data and data form.

Several characteristics should be noted in this table for each of five classes of data:

- Geology and Geophysics - Except for geological and mineralogical samples, and bottom photos, most of the raw data are digital or analog in form and are therefore suited to storage and retrieval in automated systems without extensive conversion. The derived data resulting from scientific use of the raw data are for the most part descriptive and graphical in nature.

- Marine Biology - Although not indicated, the bulk of the raw data gathered in this field is embodied in features of the gathered samples. The observations made in primary marine biology measurements are therefore descriptive and often not recorded except through preservation of samples. The principal derived data are embodied in journal articles, fisheries' reports and other hard copy artifacts.

- Marine Chemistry - Most of the raw data are digital and the derived data are descriptive, although with the increasing attention being directed at use of chemical data in corrosion studies, the design-oriented uses will lead to increased derivation of digital or graphic data.

Table II-J-3

Characteristics of Typical Classes of Oceanographic Data

Field	Primary Measurements	Raw Data Artifacts
Geology and Geophysics	Grabs, dredges, and cores	Samples
	Bottom photos	Photos
	Reflection measurements	Digital data.
	Bathymetric penetration	Analog data
	Gravity	Digital data
	Magnetic field	Digital data
	Acoustic depth recording	Digital data
	Sub-bottom seismic profile.	Analog data.
	Bottom minerology	Samples
	Seismograms	Analog data
	Acoustic attenuation	Digital data
Marine Biology	Plankton samples	Samples
	Water samples	Samples
	Bioluminescence	Digital data
	Grabs, nets, and trawls	Samples
	Scattering layer data	Digital data
	Biological sound frequency & intensity	Digital data
	Photographs	Photos
	Sonar graphs	Graphical data
	Fishing sightings	Digital data

(Continued)

Table II-J-3

Characteristics of Typical Classes of Oceanographic Data

Examples of Functional Utility	Typical Derived Data	Derived Data Form
(Geology and Geophysics)		
Geological and geophysical theoretical studies, ASW instrument calibration, undersea communication equipment development, petroleum and natural gas exploration, mineral exploration.	Stratigraphic sections	Graphic
	Geologic age determinations	Descriptive, graphical, and alphanumeric
	Tectonic studies	Descriptive
	Earth morphology studies	Descriptive
	Published articles	Descriptive, graphical, and alphanumeric
	Petroleum and mining engineering reports	Descriptive, graphical, and alphanumeric
(Marine Biology)		
Taxomonical studies, life cycle studies, flora and fauna productivity prediction, fish distribution and migration prediction.	Biota distribution charts	Graphic
	Species descriptions	Alphanumeric and descriptive
	Commercial fishing reports	Alphanumeric and descriptive
	Sonar graphs	Graphic
	Published articles	Descriptive, alphanumeric, and graphic

(Continued)

Table II-J-3 (continued)
Characteristics of Typical Classes of Oceanographic Data

Field	Primary Measurements	Raw Data Artifacts
Physical Oceanography	Geographical position and time	Digital data
	Fathometer or other depth readings	"
	Bathythermometer readings	"
	Wave and swell height	"
	Wave and swell direction	"
	Wave and swell period	"
	Tide height estimate or recording	"
	Estimated sea state	"
	Drift bottle position	"
	Subsurface current reading	"
	Shelf wave measurement	"
	Hydrologic optics	"
	Ambient noise	"
	Infrared thermometry	"

Table II-J-3 (continued)

Characteristics of Typical Classes of Oceanographic Data

Examples of Functional Utility	Typical Derived Data	Derived Data Form
(Physical Oceanography)		
Oceanographic phenomena studies, sea-state predictions for maritime use, undersea temperature data forASW systems use.	Tide and current charts	Graphic
	Ocean-atmosphere heat flow analyses	Descriptive and graphic
	Published articles	Descriptive, alpha-numeric, and graphic

Table II-J-3 (continued)

Characteristics of Typical Classes of Oceanographic Data

Field	Primary Measurements	Raw Data Artifacts
Marine Chemistry	Color and other optical properties	Digital data
	Osmometer readings	"
	Oxygen and other gas chromatographic analyses	"
	pH	"
	Inorganic analysis	"
	Organic analysis	"
	Turbidity and suspended solids analyses	"
	Sludge detection	"
	Salinity measurements	"
	Radioactivity	"
	Trace analyses (i.e., tetra ethyl lead)	"
	Conductivity and dielectric constants	"
Meteorology and Climatology	Wind speed and direction	Digital data
	Ambient temperature	"
	Hydrometer readings	"
	Hygrometer readings	"
	Cloud cover photos	Graphic data
	Barometer readings	Digital data
	Radiation count	"
	Rainfall	"
	Iceberg drift, speed, and direction	"

(Continued)

Table II-J-3 (continued)

Characteristics of Typical Classes of Oceanographic Data

Examples of Functional Utility	Typical Derived Data	Derived Data Form
(Marine Chemistry)		
Effects of chemistry on biota, corrosion studies, study of chemical property effects on optical properties, study of chemical property effects on acoustic properties, fallout effects studies, pollution buildup studies.	Published articles	Descriptive, alpha-numeric, and graphic
	Synoptic analyses of chemical/biological data	Descriptive, alpha-numeric, and graphic
	Scientific and technical reports	Descriptive, alpha-numeric, and graphic
(Meteorology and Climatology)		
Sea-surface and sub-surface temperature prediction for ASW use, current and tide prediction for weather prediction, temperature and iceberg data for climatological studies, iceberg sightings for maritime use.	Synoptic analyses of meteorological/fisheries data	Descriptive, alpha-numeric, and graphic
	Computer printouts of predicted data	Alphanumeric
	Published articles	Descriptive, alpha-numeric, and graphic
	Weather maps	Graphic
	Iceberg sightings	Alphanumeric

- Meteorology and Climatology - Most of the raw data are digital in nature, and the derived data have a wide variety of forms.

- Physical Oceanography - Most of the raw data are digital in form and the derived data are for the most part embodied in hard copy artifacts containing descriptive, graphic and alphanumerical data.

The wide variation in quality of each class or subclass of data shown in these tables results from the wide variety of sensors used in primary measurements, and the motives and modes in the use of sensor devices. To illustrate the wide variation in instruments available for primary oceanographic measurements, the following are given as examples of instruments used by the U.S. Navy for the following measurements:

o Temperature: Protected reversing thermometers, minimum thermometers, thermistors, hollow springs (bourdon tubes), infra-red (an indirect technique), and bi-metallic thermometers;

o Salinity: By titration with silver nitrate, by conductivity, by a vibrating reed densitometer, by radio frequency absorption, by hydrometer, by refractometer;

o Depth (or Pressure): Hollow springs, elastic deformation of glass (unprotected reversing thermometer), strain gauges, spring and bellows, sonar techniques;

o Current: Drift bottles and cards, propeller logs, electromagnetic logs, neutrally buoyant floats, geomagnetic electrokinetigraph;

o Waves: Wave staff with camera (or other recorder), aircraft stereo-photography, resistance-wire wave poles, capacitance-wire wave poles, accelerometers (Splashnik & Instrument ship), bottom-pressure gauges; and

o Radiation and Transparency: Eppley pyroheliometer, Secchi disc, hydrophotometers, photocell radiometers.

Another factor associated with the management of data is the wide variation in relative volume of data. Data obtained by oceanographic research institutes in the field of marine biology is of enormous volume. Scripps Oceanographic Institution has millions of fauna samples, systematically stored, identified, and embodying features for subsequent taxonomical or other use. In contrast to this large volume of data, T. J. Chow of Scripps produces as many as three pieces of data per year in his trace analysis survey of the lead tetra-ethyl content of the ocean waters. In summary, the volume and the other qualities of the data are regulated by the nature of the scientific and/or technical activity which motivates the data gathering and use activities.

3. Data Flow

Two principal classes of organizations are involved in the flow of oceanographic data: private organizations including profit-making corporations and private research institutions; and government agencies. The modes of flow are patterned after the programs and operations of these organizations, and there seem to be three primary and interrelated modes of flow:

- Data are generated or otherwise obtained in support of major missions or research programs such as the Indian Ocean Expedition and the Tropical Atlantic Expedition, and used by the research communities involved in the programs to advance knowledge of the geographic region under study;

- Data are generated or otherwise obtained in support of operational programs such as that of the U.S. Naval Operations Office, and the Bureau of Commercial Fisheries Pacific Ocean Environmental Monitoring Program; and

- Data are generated or otherwise obtained for use in engineering development projects such as in petroleum exploration within a given geographic sector and in design of submersible structures or deep diving vessels.

Data Users. Complete elaboration of all the users of oceanographic data and the modes of use is not within the scope of this study. However, examples of uses are given (from the first report of the Marine Council to the President) to illustrate the modes of use and the primary sources utilized:

"—the scientist who is interested in the phenomenology of the oceans for scientific objectives but whose knowledge and perception are the basis for a rigorous understanding of the oceans and atmosphere;

—the naval planner concerned with antisubmarine warfare who must understand undersea phenomena that aid concealment;

—the climatologist who must acquire and analyze large quantities of often seemingly unrelated information in order to understand local, regional, and world climate;

—the meteorologist, oceanographer, and seismologist who are concerned with the influence of the oceans on the weather over ocean and land areas and who must warn of hurricane, storm surge, and of tsunami sea waves of destructive character;

—industrial managers undertaking expensive offshore mining or oil-drilling operations who need information on the ocean bed and water conditions above it; and

—the commercial or sport fisherman who will be able to draw on oceanic data, and aircraft- or spacecraft-derived surveillance, to predict location and density of fish stocks."

To illustrate further, the following description of the utility of oceanographic data to Naval operations is given (from Naval Training Device Center Document 1494-1):

o "Geophysics and Physical Oceanography

The scientific study of marine physics produces beneficial results in the solution of problems and situations in virtually all sectors of naval operations. Foremost in the conduct of surface operations is knowledge of the surface at present and in the immediate future. Navigation requires knowledge of currents, earth's magnetism, and air movement. (Meteorology is a specialized branch of physics.)

Waves influence any surface movement of vehicles and the men, material, and ordnance transported. Surface condition is pre-eminently important in ship control, replenishment, and amphibious assault. Sea turbulence also influences air operations, submarine detection, mine laying, mine surveying, and reconnaissance.

Temperature is a determining factor in sound velocity, and, consequently, in submarine detection. Temperature is an important influence on the rate-of-fouling and corrosion affecting vessels and structures immersed in the sea.

Density determines velocity of sound transmission. Effective submarine operation requires a thorough knowledge of density conditions, and these are determined by knowledge of temperature, salinity, and pressure (depth).

Geomagnetic forces include gravitational fields affecting surface navigation and ordnance operation. In certain instances, mine warfare is predicated upon particular locations of the force field.

o "Chemistry

Chemical characteristics of sea water influence corrosion control of ship hulls, screws, and exposed sessile equipment arrays. Salinity is also a determining factor of sound velocity in water, and sound is the most important available avenue for detecting submerged objects, either mobile or stationary.

o "Biology

Uncontrolled organic populations causing bottom fouling seriously affect surface operations. Fouling creates potential adverse effects of submarine search

by surface vessels when a false, biological target continually causes loss of time and effort, expenditure of ammunition, and inhibition of effective search tactics.

Fouling presents a major problem in the maintenance of subsurface arrays, in the effective operation of sound-generating sources, and the preservation of metallic surfaces. Corrosion may be associated with fouling, as the organisms may attack or clean vulnerable surfaces. Although fouling occurs at all depths, it is particularly severe in the upper 400 feet of water, in warm waters, and on the bottom. In order to combat effectively problems of fouling and boring organisms, it is necessary to understand the ecological relationships of Benthic and planktonic populations.

Bioacoustics, the study of soniferous species, is important whenever sound discrimination is essential in identification/classification of targets. Benthic populations, as well as nekton, may contribute to the sound environment to such a degree as to disguise transmission by sonar. These noises can arrest the effectiveness of passive listening devices, such as "heralds" used in harbor defense or sonobuoys used in submarine search. In part, the welfare of underwater reconnaissance swimmers is dependent upon the frequency of occurrence of noxious or pestiferous species. Among such species can be found vicious creatures (sharks, cowries, and morays) as well as those of somewhat lesser hazard (sea snakes, stinging urchins, certain jellyfishes, and some corals).

o "Geology

The shape and depth of ocean basins, the land forms that surround those basins, and the discrete components of the ocean bed have impacts on naval operations.

Submarine detection equipment cannot be emplanted on the ocean floor nor installed in the vicinity of harbors unless prior consideration is given to geological configurations. Information on bearing

strength of sediments, rate-of-sound through bottom deposits, and occurrence of rocky reflecting surfaces or sound-absorbing muds is necessary for efficient emplacement.

Coastal land forms vary around the world. As it is seldom possible to select beaches for amphibious assault, each of fifteen fundamental land forms bordering the seas must be examined for the highly important ramifications in devising landing assault techniques. The features should be studied by students of warfare.

In mine warfare, it is essential that information be available concerning the nature of the bottom. Is it muddy enough to cause acoustic mines to sink from sight? Are the bottom deposits of sand, silt, ooze, or rock; and which will support a minecase anchor? Is the bottom rent with crevices or ripples into which currents will cause the mine to "walk?" Does bottom vegetation grow abundantly enough to contribute seriously to fouling? This, and other questions, should be illustrated for the student.

- "<u>Meteorology</u>

  In amphibious operations prior climatic intelligence is necessary to assess the need for support from other naval elements (e. g. , logistics, submarine defense, mining, or air cover).

  Air operations are severely restricted by adverse weather conditions such as icing, fog, or wind storms. Ship control is as easy in gentle weather as it is difficult in foul, and mine sweeping effectiveness is largely dependent upon sea conditions. "

The utility of these classes of data in Naval Operations is summarized in Table II-J-4 to illustrate specific divisions of interest in this operation. Another example of government operational use of oceanographic data is in efficient support of commercial fisheries operations. Major strides have been made toward definition of the specific data requirements of fisheries by the Bureau of Commercial Fisheries of the U.S. Department of the Interior. The following excerpts from a BCF report on the Pacific Ocean Environmental Monitoring Requirements indicate progress toward definitions of fishery requirements and efforts to meet them:

Table II-J-4

Utility of Oceanographic Data to Naval Operations

CLASS OF DATA	SURFACE			SUBSURFACE			AMPHIBIOUS		OTHER OPERATIONS		AIR		
	Effects on Vehicles	Effect on Search Detection & Target Identification	Effect on Weapon Delivery	Effect on Vehicles	Effect on Search Detection & Target Identification	Effect on Weapon Delivery	Reconnaissance	Assault	Mine Warfare	Harbor Defense	Effect on Vehicles	Effect on Search Detection & Target Identification	Effect on Weapon Delivery
Marine Biology													
False Targets		X			X							X	
Fouling	X	X		X	X				X	X	X	X	
Littoral Vegetation							X	X					
Noxious Species							X	X					
Bioacoustics		X	X		X	X			X	X		X	X
Bioluminescence		X							X			X	
Marine Chemistry													
Corrosion	X			X					X	X	X		
Salinity		X		X	X				X	X		X	
Precipitation					X		X						
Solutes	X												
Electrolytes	X			X					X				
Geology													
Basin Structure		X		X	X					X		X	
Topography		X		X	X		X	X		X		X	
Beaches							X	X					
Sediments	X	X			X		X			X		X	
Geodesy				X						X	X		
Geophysics and Physical Oceanography													
Temperature		X		X	X			X	X	X		X	
Density		X		X	X							X	
Pressure		X	X	X	X	X						X	X
Geomagnetics	X			X					X		X	X	
Waves	X			X	X		X	X	X	X	X	X	X
Currents	X			X			X	X	X	X		X	
Sound			X	X	X	X			X	X		X	X
Ice			X	X	X	X		X	X	X	X		X
Underwater Visibility							X			X		X	
Meteorology													
Precipitation					X			X			X	X	X
Visibility							X	X	X		X	X	X
Winds								X	X		X		X
Air Temperature					X			X			X	X	

"The Bureau's oceanographic program, outlined in Oceanographic Program for the Bureau of Commercial Fisheries, August 17, 1965, is designed to answer four principal questions, as follows:

(1) What are the kinds, geographic distributions, and abundance of resource organisms, and what are the characteristic environmental conditions of each?

(2) How and why do populations of marine resource organisms vary from time to time in their distribution, abundance, and availability to harvest by man and as food for other organisms?

(3) How can the efficiency of fishing gear and fishing operations be increased?

(4) What are the possible ways in which the marine environments may be altered to enhance the productivity of useful organisms?

"To answer these questions in their entirety requires time-series of biological, oceanographic, and meteorological data. Relationships discovered between the organism and the environment are often empirically derived and frequently lead to little understanding of oceanographic and biological processes involved. The availability of time-series data and analyses thereof lead to increased understanding of dynamics of ocean change which may further increase understanding of why marine organisms vary from time to time in their distribution, abundance, and availability "

This report continues with a description of a plan for programmed generation of specific types of data in support of Pacific fisheries operations, and collection of existing data that would support the program. An important characteristic of oceanographic data that emerges from this example is its enlarged value when used synoptically, i. e., when several types of simultaneous measurements are correlated. In the particular instance of fishery

operations, synoptic studies of the biospherical behavior and environmental conditions facilitate predictions of fish resource distribution and movement. Notable among the resources which this program will use are the data available from the National Oceanographic Data Center in Washington, D. C., and the Fleet Numerical Data Center at Monterey, California.

The data used in phenomenological investigations undertaken by private research institutions are of a different quality than those used in support of operations such as the two examples given in the foregoing paragraphs. The full span of biological, geological and geophysical, chemical, meteorological and physical oceanography data is used by individuals or groups investigating phenomena of a specific class (i.e., plankton studies) for a given region in the oceans. The user communities are tied by close personal communication linkages either within a single institution, or one or two institutions which focus on common regions. For example, Woods Hole Oceanographic Institution and the Lamont Geological Observatory are concerned with phenomena in the Atlantic and Caribbean, while Scripps Oceanographic Institution and the University of Oregon are concerned with the Pacific Ocean. The following are types of data gathered by scientists on board a specific research cruise sponsored by Oregon State University:

<u>Cruise</u>: R/V Yaquina (April-July, 1967)

<u>Physical Oceanography</u>

solar radiation effects
subsurface current
sea level changes & upwelling & tides
heat storage
shelf waves
seiching (oscillation of landlocked water)
hydrography
hydrologic optics

Geology (samples gathered)

estuarine sediments
estuarine ecology
shelf sediments
silt minerology (deep sea)

Geophysics

gravity
seismology
magnetics
thermal effects

Chemistry

reactions
oxygen-phosphate
pH
dissolved N_2 & CO_2 by gas chromotography

Radiochemistry

dissolved organics
trace element analysis
radioecology of benthos fauna

Biology

amphipods
deep scattering layer data
fauna ecology
upwelling & biomass
benthic fauna ecology
echinoderm productivity
deep sea fish systematics
deep sea fouling
phytoplankton pigments, physiology, ecology,
microbiology
coastal invertebrate

Samples and data from such research cruises are stored within the specific departments of the institutions and used as required for the specific research projects of the scientific staff. Geological photos and samples are analyzed as required for scientific investigations and biological samples are preserved and similarly archived for future use. Physical oceanographic data are stored in the scientists' laboratories and offices, and where possible within the financial constraints of research funding (largely by the Navy Office of Naval Research) are transmitted to the National Oceanographic Data Center. Geological samples are also sent to the Smithsonian Institution Oceanographic Sorting Center, where they are classified and cataloged. Generally speaking, the researchers use data resources available within their own community of interest including the samples and publications to which they are readily exposed. Meteorological data collected in cruises are utilized to account for anomalies in geophysical or other data which might have arisen from the effects of the environment on the primary sensing operations. These data are generally archived in cruise logs to which all data are generally associated. Some oceanographic institutions, such as Woods Hole, catalog and centrally archive computer programs used in the manipulation of raw digital data.

Data used for the engineering development of oceanographic systems, structures, and vehicles include a wide variety of materials, electronics, acoustics, and structural design information. The user patterns are closely allied to those which are seen in the aerospace and other mission-oriented engineering development fields. Most systems development activities rely heavily on prototype development and testing for proof of performance capability, and the data flow is generally limited to a single project effort, such as construction of Woods Hole's submersible vehicle, the Alvin.

Data Generation. The principal modes of data generation in the field of oceanography are closely allied to the user patterns. In the phenomenological research sector, the users and/or user communities generate their own data to suit their own quality requirements. Operational data, such as fishery data, come from a wide variety of

sources including government-operated services and research institutions. The following abbreviated summary lists typical data generators and the classes of data they generate:

DATA GENERATOR/DISSEMINATORS

Federal Government	Data Classification
Atomic Energy Commission	radiation level
Coast & Geodetic Survey	bottom topography
	gravitational & magnetic fields-tides
	ocean currents
	temperature
	ocean density
Geological Survey	continental shelf topography
	geological & geophysical mapping
Bureau of Commercial Fisheries	marine ecology
	currents
	temperature
Bureau of Mines	geology & sea mineral resources
Public Health Service, Office of Saline Water	water resources data
Army Corps of Engineers	water resources data
Navy Bureau of Yards & Docks	coastal engineering data
U.S. Maritime Commission	surface vessel design data
U.S. Navy	
Special Projects Office	surface and submersible vessel design data
Ship System Command	
Air Systems Command	
Naval Research Lab	
U.S. Coast Guard, Ice Patrol	offshore platform & dock design
	iceberg data
Weather Bureau	ocean currents
	tidal data
National Oceanographic Data Center	full spectrum of oceanographic and related data on an international scale
ASWEPS (Antisubmarine Warfare Environmental Prediction System)	weather & safety data

Federal Government (cont'd)	Data Classification
AUTEC (Atlantic Undersea Test & Evaluation Council	ocean floor topography physical oceanography (acoustics)
Smithsonian Institution	geological and biological samples
International Programs/Institutions	
Indian Ocean Expedition, Tropical Atlantic Expedition	full spectrum of data for target region of research
International Ice Patrol	iceberg data
Soviet Interdepartmental Coordination Scientific Council	full spectrum of data types
Oceanographic Society of Japan	biological resources (fisheries) ship safety data
Liverpool University - Tidal Observatory	ocean floor topography
Cambridge University Dept. Geodesy & Geophysics	marine geology
Canadian Fisheries Research Board	biological resources, fisheries data
Ocean-Wide Survey Program	international program to coordinate data flow
IGY-Data Center A	
-Glaciology	glaciology - iceberg
-Oceanography	physical oceanography
Intergovernmental Oceanographic Commission	worldwide cooperation (does not handle data)
Private Institutions	
American Geographic Society	continental shelf data physical oceanography marine biology geology geophysics
Rhode Island University	fishery resources marine biology limnology chemical oceanography

Private Institutions (cont'd)	Data Classification
Arctic Institute of North America	iceberg data
Scripps Bathythermographic Data Processing & Analysis /Ocean Data Archives	physical oceanography chemical oceanography climatic correlator
Scripps Marine Geology & Geodesy Department	deep core sediments
University of California Water Resources Dept.	water resources
University of Oklahoma Bureau of Water Resources	water resources
Lamont Geological Observatory	geology physical oceanography
Department of Oceanography & Meteorology, Texas A&M (see IGY World Data Center A)	geology physical oceanography
Woods Hole Oceanographic Institution	marine biology geology & geodesy physical oceanography
Institute of Marine Science University of Miami	basic marine biology
Department of Oceanography University of Washington	physical oceanography chemical oceanography
Polar Studies Institute	glaciology

Two significant recent developments that will influence the availability of oceanographic data are the use of satellites and the development of a national ocean data system using unmanned buoys. In an address to the Fifth Symposium on Remote Sensing of Environment given in April, 1968, by Homer Newell of NASA, the following prospects for use of satellites were set forth: "The large-scale features of the oceans are dynamic and can only be monitored adequately with frequent repetitive measurements over wide areas.

Because most of the oceans are never seen by man, oceanography should lend itself ideally to remote sensing by satellites. The oceanographic features studied and developed to the point of satellite applicability are sea surface temperature and currents, sea state, sea ice, and marine life detection. It would be most desirable to make these observations at frequencies which can penetrate the earth's incessant cloud cover.

"Sea surface temperature gradients and discontinuities have been studied from aircraft in the visible, infrared and microwave regions of the spectrum. The visible and infrared regions are available for sensing only during cloud-free conditions, the infrared is useful at night as well as day, and microwave is useful under all conditions, including cloud cover.

"Nimbus II high resolution IR images of sea state from 1150 km altitude have been obtained, and computer gray scale plots of temperature contrasts have been made. Sea surface temperatures have also been inferred from cloud patterns in high altitude (22,000 miles) ATS-1 imagery. The recent color photographs of the earth taken in the Apollo 501 mission, near apogee of 9,000 miles, are of particular use for sea surface study and have been interpreted successfully, largely due to their quality which is superior to that of the imagery from the meteorological satellites.

"Scientists have long been searching for a method to measure sea state in all kinds of weather on an oceanwide basis as an aid to the shipping industry and for weather forecasting. It is common practice to infer sea state conditions from wind reports. One method of measuring sea state is the analysis of wave patterns and sun glitter in aircraft photographs of the sea surface. New techniques utilizing passive microwave and radar reflectance measurements are currently the most promising for sea state determination from high altitude since they are sensitive to wave characteristics and can be made with no appreciable attenuation in the presence of storms and clouds. Investigators have shown recently that by plotting reflected radar energy against the angle of incidence at the sea surface, one obtains well-separated signatures for various states of sea roughness. The data were substantiated by MSC aircraft measurements in late

1967 off Newfoundland, and further tests are currently in progress off the coast of Iceland. . . . All in all, the field of oceanography shows considerable progress in the application of remote sensing techniques. Technologies for surveying the ocean temperature and sea state condition are probably the nearest at hand of all the earth resources observations. Consequently, we may expect that these will be the first to reach operational status."

The other prospective new source for oceanographic data is a network of data-gathering buoys being developed by the Navy's Office of Naval Research. According to the buoy-network plan, data will be generated by unmanned buoys and telemetered to data centers. One buoy has been built and tested, and plans are being completed for more.

4. Representative Problems

The principal problems associated with the field of oceanography are related to the needs of two communities of data users:

- The Phenomenological Research Community. Three problems plague the oceanographic researcher in the environment of increasing organization of programs and data management activities. First, inadequate support is being provided to facilitate the completion of data formatting requirements of centers such as the National Oceanographic Data Center (NODC). If funds were provided, specialized personnel would be trained to take over the data formatting functions so that scientists would not have to perform functions such as conversion of analog bathythermographic data and bottom photo data to digital forms required for NODC compatibility. Estimates of funding requirements have not been made, but the order of magnitude is the amount of funding presently required to finance the actual research activities. The Office of Naval Research

does not provide funding for data management apart from those activities within the context of scientific activity. Unless this situation is changed, the massive oceanographic data resource, generated by the research community, will not become completely available to other oceanographic data users such as the Naval operations community.

The second problem which plagues the research community is the mismatch of the data quality required for the performance of research functions with the quality of the data available from evolving data centers such as NODC. Scientists at several research establishments have stated that data generated by inadequately trained Navy recruits using nonstandardized instruments will not suffice. Part of the answer to this problem may result from the establishment of a National Oceanographic Instrument Center, and the proper coding of data stored in data banks to identify instruments and environmental conditions.

The third problem associated with oceanographic research data has been that data gathered during cruises could not be adequately evaluated until the cruise had ended and the scientists returned to their land-based laboratories and computer facilities. Often data gathered during cruises for a critical geographical position is found to be erroneous or anomalous due to environmental conditions or instrument failure. Resolution of this problem may result from promising results of pilot tests of a shipboard computer at Woods Hole Oceanographic Institution. According to C. O. Bowin of Woods Hole (Proceedings of the Fourth Naval Symposium on Military Oceanography, 1:253-264), it is now possible to chart magnetic field, Bouger anomalies and other data while a

cruise is in progress, so that any errors or faults in collected data may be corrected while the research vessel is still near the geographic position in question. Further funding for such developments will assist in producing higher quality data, and the possibility of automating the formatting of data to meet NODC and other data center requirements. Automation and standardization of data forms and formats will facilitate synoptic studies for research and operations use as well.

- The Operations Management and Support Community. The requirements of operations communities for up-to-date and complete data relative to their operation are not being met. Lack of critical evaluation, inadequate identification of data sources, archaic methods of filing, and lack of standardization of data form and format are the principal underlying problems. Before the data requirements of operational communities can be satisfied, study of the exact requirements for each mission of each community must be defined, and data gathering and handling missions must be formulated. This approach is in keeping with the successful tradition of managing scientific and technical activities aimed at attainment of a specific goal or objective. It is hoped that the SDC study of oceanographic data underway at this writing will identify the primary goals and missions.

Selected Bibliography

Effective Use of the Sea, Report of the Panel on Oceanography, President's Science Advisory Committee, June 1966.

Introduction to the National Oceanographic Data Center, NODC Publication G-1.

Marine Science Affairs - A Year of Plans and Progress, The Second Report of the President to the Congress on Marine Science and Engineering Development, March 1968.

Marine Science Affairs - A Year of Transition, The First Report of the President to the Congress on Marine Science and Engineering Development, February 1967.

Oceanographic Information Sources, National Academy of Sciences Publication No. 1417.

Oceanographic Reports, U. S. Coast Guard Reports No. 8, 11, and 13, 1965.

Oceanography 1966, National Academy of Sciences Publication No. 1492.

Scientific and Technical Personnel in Oceanography, Interagency Committee on Oceanography Pamphlet No. 21.

Transactions, American Geophysical Union, Vol. 49, No. 1, March 1968.

III CURRENT STATUS OF BASIC, DEVELOPMENTAL AND APPLICATIONS DATA ACTIVITIES

A. Introduction

One of the more significant overall findings that resulted from this survey/analysis is the clear evidence of distinct roles and modes of data flow for the research, developmental and applications phases of the science and technology cycle. These distinctions are important in the establishment of plans and programs for the management of scientific and technical data. The descriptions of principal data activities in 10 selected fields of science and technology, to a certain degree, indicate these distinct roles and modes of data flow. In this section, data and data flow are first characterized within the framework of the research, developmental, and applications phases of the science and technology cycle. Then an analysis of the problems associated with each phase is presented, and finally, an outlook on future data management practice within each phase is projected.

1. Definitions

To set the analysis framework, definitions of basic, developmental and applications activities are first set forth:

(1) Basic data* are those which are generated by research investigations seeking to describe phenomena, substances, and systems;

(2) Developmental data are those which are generated as the result of practical application of scientific knowledge, materials or techniques to meet a technical need; and .

(3) Applications data are those which are used in the production, operation and maintenance of equipment, material products, and operating systems of all types.

---

* Basic data are frequently scientific data.

These classes of data are closely associated with the activities which tend to use or generate them. The three corresponding classes of data activities are as follows:

- Discipline-Research Data Activities, which are directed at the generation, management, and handling of data to increase knowledge and the understanding of phenomena. The motivations of the scientist are to proceed from more empirical to more rational measurement strategies, to reconcile his measurements with relatable measurements, to generate broader rationalizations, and to produce documentation displaying the linkage between primary measurements and the highest level of generalizations for which they are valid.
- Mission-Developmental Data Activities, which are associated with practical use of scientific and/or technical knowledge for the solution of a technical problem or the attainment of a scientific or technological goal. In the course of performing these activities, scientists and technologists attempt to utilize proven theories, rationalizations of a very high order, and models of phenomena to predict the data which they will require. But in many instances, rationalizations, theories, and models do not exist, and the scientist or technologist must resort to empirical correlations or his own primary measurements to obtain the required data.
- Applications-Product Data Activities, which are associated with the application of both basic and developmental data for the production, operation, and maintenance of products and operating systems of all types, as well as the performance of scientific and technical services.

2. Scope

According to J. W. Carlson, Senior Staff Economist of the President's Council of Economic Advisors, the United States Federal Government is the largest single spender in the international research and development community. In the fiscal year 1967, nearly 15 billion dollars were spent by the Federal Government on research and development activities, nearly 63 percent of the 24 billion dollar national total.* The U. S. investment in research and development since World War II totals well over 100 billion dollars.

The generation, handling, and application of scientific and technical data are principal elements of all scientific and technological activity. As yet, no reliable estimate has been made of the percentage of research and development expenditures which are directed to data activities. However, it has been estimated that scientists and engineers spend anywhere from 20 percent to 30 percent of their working time acquiring data. A reasonable estimate of the amount of Federal money currently being spent for just one facet of the entire data handling process -- that of data gathering -- is, therefore, approximately three billion dollars annually. To date, attempts to obtain precise totals for the costs of scientific and technical activity have been unsuccessful because of an inability to separate these costs from the costs of other functions involved in performance of scientific or technical work.

As the missions assigned science and technology have become broader and oriented more to solution of technological and sociological problems rather than to the production of materials and products, the responsibilities of funding and managing scientific and technical activities have shifted more and more to the Federal Government. For the most part, the shift in assumption of responsibility from industry to Federal Government for management of scientific and technological programs and supporting data activities has been strongest in the fields for which the profit motive has not been adequate to attract industry investments. A good example is the collection of general purpose scientific data. A recent National Science Foundation study in this area indicated that Federal

*New York Times, 14 August 1966, pp. 10ff.

obligations for collection of data of general scientific use amounted to 412 million dollars in fiscal year 1968, and that these expenditures have grown at the rate of eleven percent per year since 1962. Of these expenditures (averaged over a five-year period), 70 percent were for data describing natural phenomena and 30 percent for data describing social phenomena. Figure III-A-1 displays the pattern of Federal expenditure and the distribution thereof for the fiscal period from 1962 through 1967. Ninety percent of these total expenditures were managed by the Departments of Commerce, Defense, Interior, Agriculture, Labor, and Health, Education and Welfare. Within these departments, the following were the largest investors in scientific and technical data:

Department	Agency	Principal Field of Science
Commerce	Environmental Science Services Administration	Environmental and Geosciences, Oceanography
Defense	Navy Department	Environmental and Geosciences, Oceanography, Weapon Systems Technology, etc.
Interior	Geological Survey	Geosciences
Agriculture	Soil Conservation Survey, Statistical Supporting Services	Agricultural Sciences
Labor	Bureau of Labor Statistics	Social and Political Sciences
HEW	Public Health Service	Biomedical Sciences

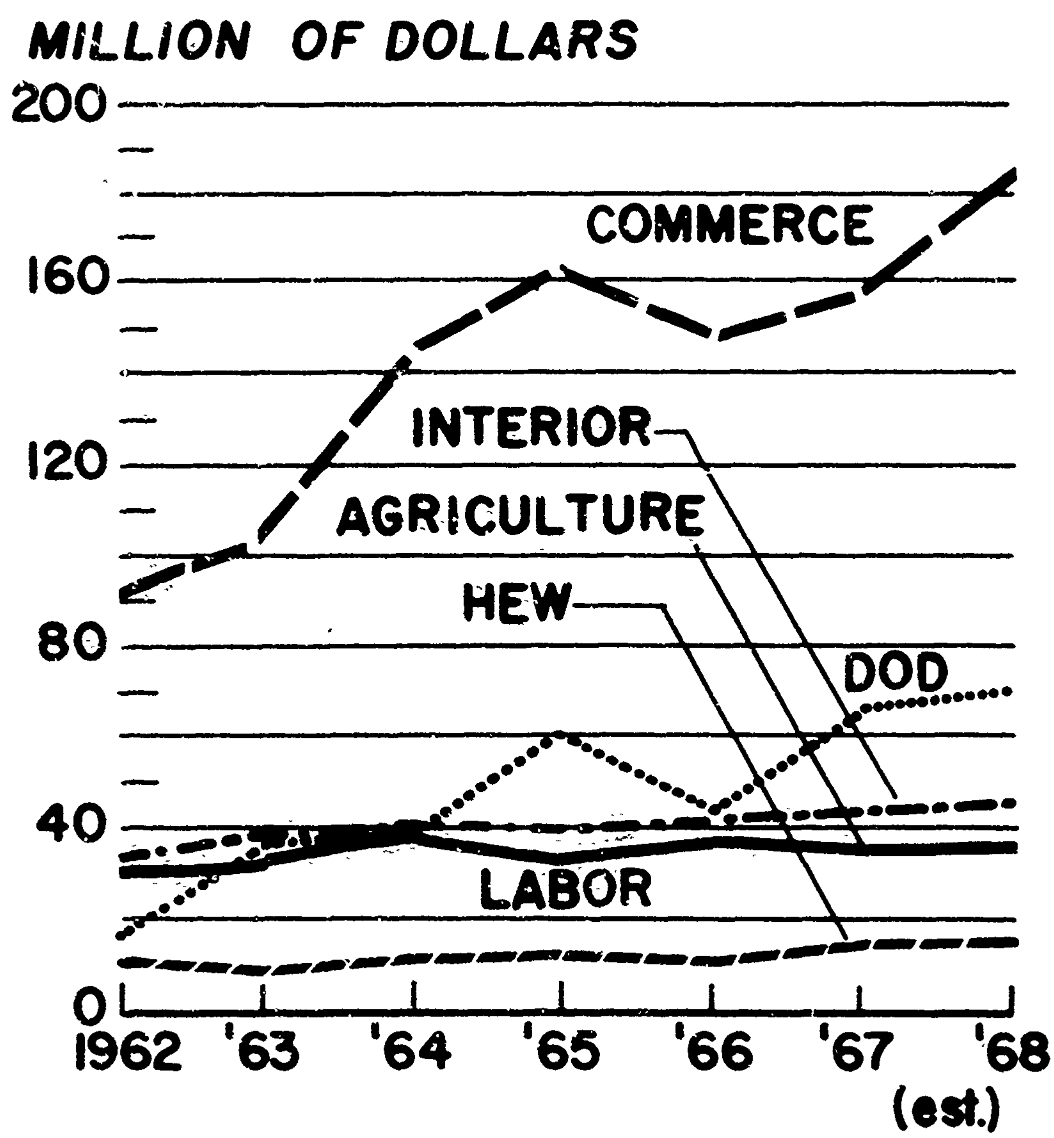


Figure III-A-1 Distribution of Federal Support of General-Purpose Data Generation

(Source: National Science Foundation)

According to the NSF source, the data gathered in these Federally sponsored activities are distinct from other scientific and technical data in two respects:

- The direct benefits derived from their application are not sufficiently definable so that industry could market them, and therefore industry is not motivated to generate, store, retrieve or disseminate them; and

- They are of such national significance that the Federal government is obliged to perform all or most of the generation, storage, retrieval, and subsequent dissemination functions.

In this study, a further generalization is revealed concerning these data. For the most part, the data generated under the support of partial or total Federal funding are those which pertain to the fields which are, first, of high national significance, and secondly, associated with the scientific and technical fields for which inadequate rationalization, theory or mathematical models exist to facilitate reliable data prediction. Note that the emphasis is on the environmental and geosciences, and to a certain extent, the biomedical and social sciences, in which fields the level of rationalization has not reached that in the physical sciences and engineering. In the latter fields, it has been estimated that more than 50 percent of the data required to meet both industry and government needs may be predicted using theories, models, or mathematical rationalizations.

This perspective, while obvious to most viewers of the national scientific and technical effort, is one that is absolutely essential in formulating policy and plans relative to data management coordination and control in the United States. Both the financial and technical scope of government and industry functions and roles in the achievement of efficient and viable data management must rest on a foundation of need assessment. It appears

from this preliminary survey/analysis of nationally significant scientific and technical data activities, that the principal obligations of industry and government have been viewed (although inadvertently) in the past, as follows:

- Industry has tended to satisfy its own requirements for directly applicable basic data, most developmental data, and all applications data relevant to industrial products and services; and

- Federal Government has performed or supported data activities which generated, stored, retrieved, processed, and disseminated data which industry could not viably support. These activities were associated with basic, developmental, and applications data, but not associated with consumer and/or industrial capital goods and/or services.

B. Data Activity Characterization

1. Overall Data Characterization

Each of the basic, developmental and applications classes of data has unique characteristics resulting from the quality and quantity requirements of the scientific and technical activities with which it is associated.

- Basic Data are generated to establish or verify a theory or rationalization of some phenomenon. Therefore, the data tend to be raw, and their quality becomes increasingly refined as the objective of rationalization or theory verification is achieved. This process is illustrated in Figure III-B-1. In this figure, the progress toward development of systematic expositions and/or codified data compilation is shown. This study of ten selected fields of data activities in science and technology indicated that basic data development in the physical sciences and engineering is further advanced toward rationalizations than in the social, life, and geosciences. As the result of this quality feature,

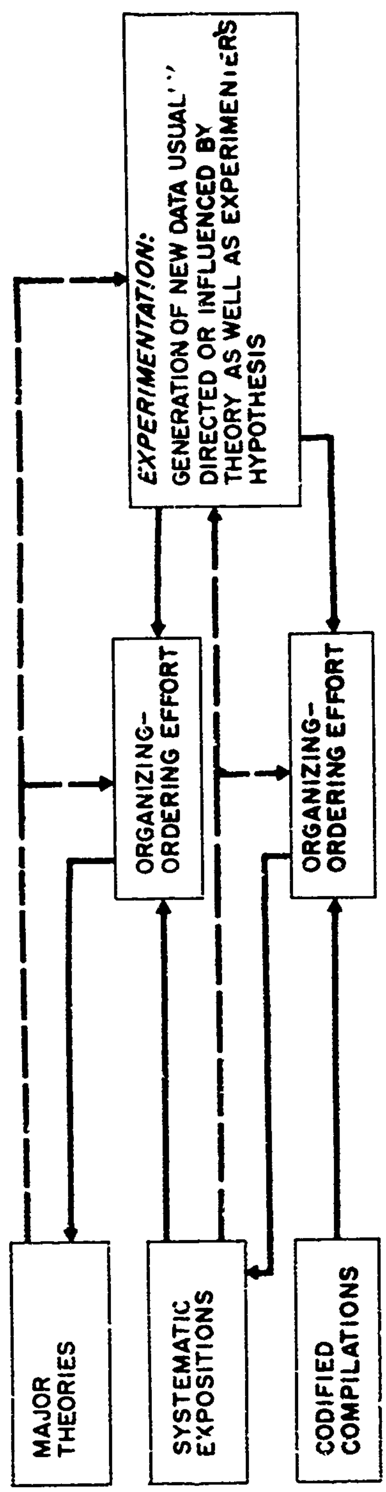


Figure III-B-1 Discipline-Research Data Flow

experimental and data evaluation or ordering activities are far less sophisticated in these latter three sciences and the data quality is much more raw and voluminous. For example, in the field of environmental sciences, weather prediction must be based on the collection of very large volumes of data, because no adequate model of the environmental system enables prediction of environmental phenomena from limited numbers of data points.

- Developmental Data, by their very nature, are far more accurate since they serve as the basis for accomplishment of scientific or technical missions often involving large investments. In many instances, the state of advancement of first principles does not permit computation and prediction of required data using rationales or models, so originally produced data must be used to satisfy the accuracy and precision requirements of the scientific or technical function. Figure III-B-2 illustrates the data use pattern that regulates the quality of developmental data. Where possible, major theories, systematic expositions or, preferably, codified compilations are used to generate data of adequate accuracy, precision and nature to satisfy mission or project requirements. Because of the quality requirements of most developmental activities, whether they involve an Apollo Project or design of a cyclotron, the data volume is specifically controlled by the scope and nature of the associated mission.
- Applications Data are highly evaluated and have a very high precision and accuracy, although end use may determine that the quality requirements are less for some applications than for others. For example, vendor and training data associated with equipment or systems usage are of a very high precision and accuracy, because they are routinely and unquestionably applied in activities for which safety and efficiency depend on their quality, while data used in product promotion are of lesser quality. This rule applies whether application data are applied in discipline-research, mission-developmental, or applications data activities.

2. Data Packaging

The modes of data use and traditions of information management for the most part have determined the mode of data packaging for basic, developmental and applications data.

- Basic Data Packaging. According to Harvey Brooks, in his paper that is included in a report by the National Academy of

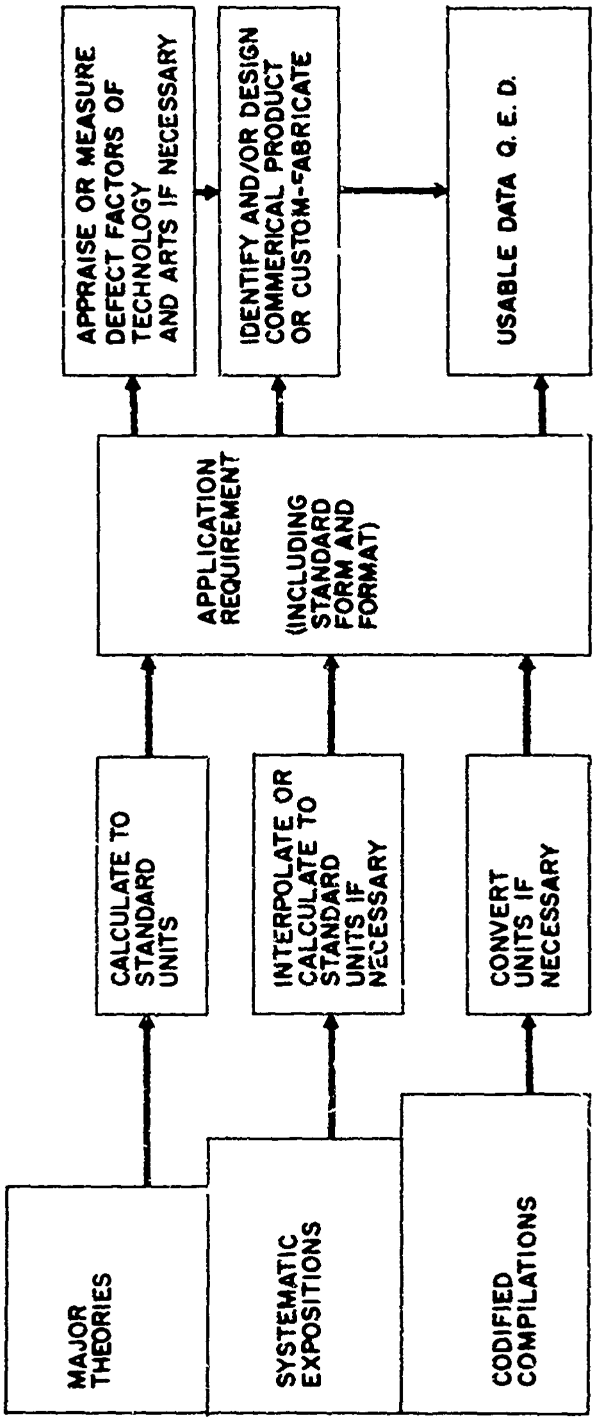


Figure III-B-2 Mission-Developmental Data Flow

Sciences to the House Committee on Science and Astronautics (Applied Science and Technological Progress, p. 38-39, 1967) "Numerous observers have commented upon the differences between the communication systems within science and those within technology. Science has an elaborate system of public documentation with strong sanctions operating on the individual scientist to make full use of and give proper credit for previous work relevant to his own ... Within technology the communication pattern tends to be more localized and more confined to operational channels. One reason for this, of course, is that much technological innovation is harder to verbalize and to document. The intuitive aspect of invention,....., makes it more dependent on face to face contact and learning by doing."

In summary, most of the basic data formally communicated are packaged within the context of the traditional basic journal publishing activity. This activity is described by the chart shown on page 129. The quality of data thus packaged is controlled by the publication editorial boards, the motives of the authors, the authors' supporting institution and the sponsoring agency.

- Developmental Data Packaging. A wide variety of data packages is used in formatting the developmental data. The following brief table indicates some examples of typical data packaging formats used.

Type of Data	Typical Data Packages	
	For awareness	For problem solving
Uniquely Applied Data	Meetings Consultants Research reports Internal memos Supplier personnel	Specifications Handbooks & manuals Standards Drawings Computer programs Product bulletins Test reports
Routinely Applied Data	Trade publications Journals Meetings Texts Supplier personnel Manual revisions	Specifications Handbooks & manuals Product bulletins Catalogs Trade publications Computer programs Test reports Data compilations

Packaging of developmental data is an art generally adapted to meet the requirements of specific missions. The chart on page 16 shows the spectrum of data packaging items used in typical aerospace missions or projects. Smaller-scale missions or projects such as development of a pharmaceutical product or manufacturing process involve far less complex packaging arrays.

- Applications Data Packaging. The traditional packages for applications data which are used in scientific and technical operations, are functionally designed to facilitate specific functions:

  (1) Training data are contained in artifacts that aim at bridging the gaps between existing and desired skills, knowledge and attitudes; experience has proven that within the present state of the training art, the most useful artifacts are hard-copy programmed instruction or straight text, instruction sheets, motion-picture film and film strip, flip cards; and
  (2) Vendor data are contained in catalogs, data sheets, advertisements, and operating manuals; vendor data used for product promotion as opposed to support of product use are of a lower detail and quality level and are formatted more for attitude modification than for knowledge enhancement.

3. Data Flow

The flow of basic, developmental, or applications data is interrelated with the data packaging modes used. . As mentioned earlier, the traditional mode of basic data flow is via the publication channels. Developmental data flow is more complex and therefore, far more difficult to describe because of the wide diversity of technical functions that are performed within the framework of mission-developmental data activities. Figure III-B-3 illustrates the communicable data that encompass the plans, status, progress or results of research and development activities. The data flow around the research, development, testing and engineering cycle starting from user reports, research activities or advanced development progresses clockwise toward product development; and the data flow proceeds concurrently via an array of data-containing

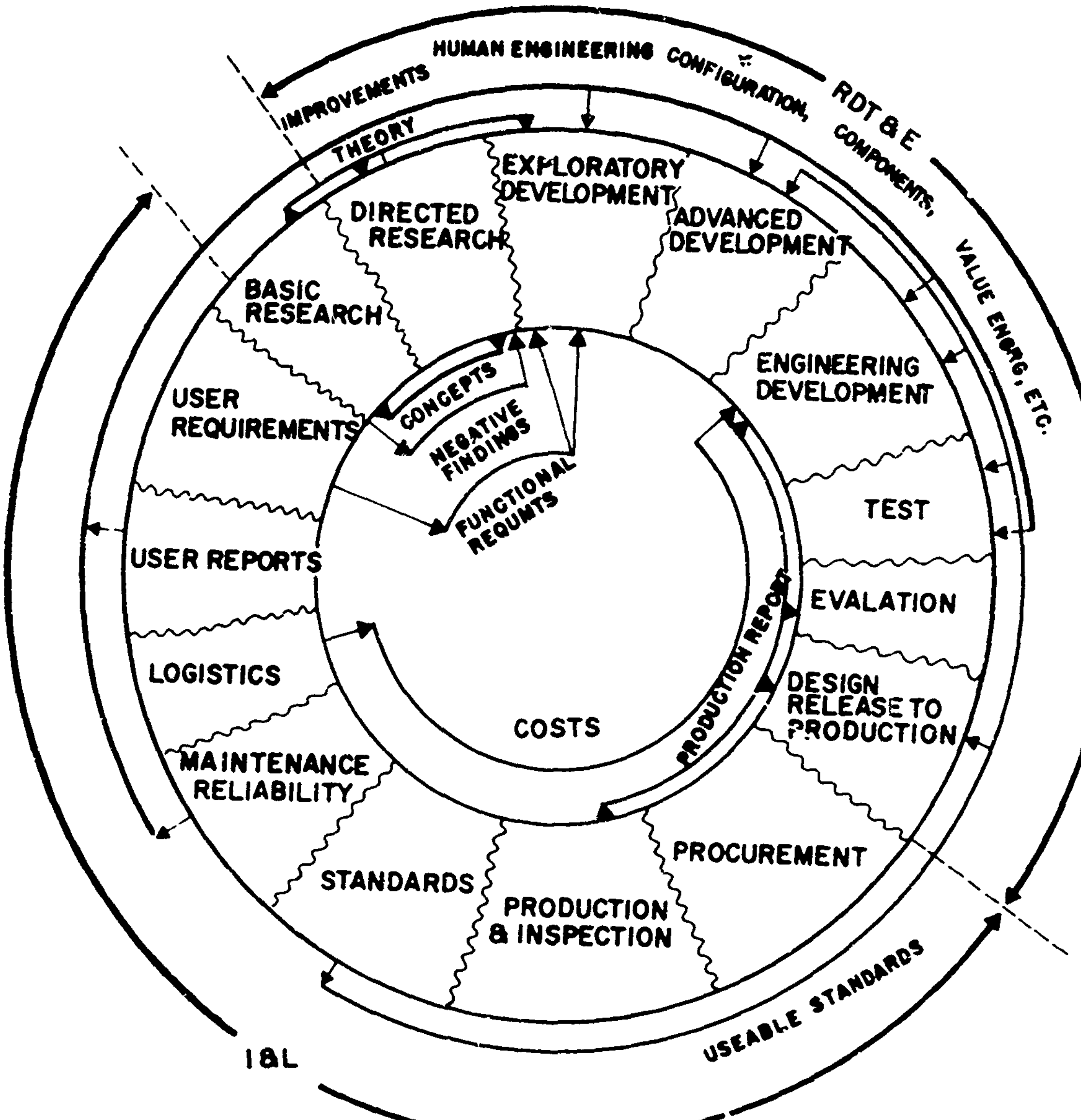


Figure III-B-3 Developmental Data Flow Pattern*

*Source: Directory of Army Technical Information. Engineering Data and Information System (EDIS) Concept and Action Plan Report. Report #EDIS-1. By S. A. Goldber, et. al. July 1964. 37 pp. AD 444700L.

artifacts which are specifically designed to meet specific data usage requirements along the way. Figure III-B-3 illustrates that basic or discipline-research data evolving out of, and used in basic or directed research interface primarily with the early development stages and, to a far lesser degree, with later development stages.

- Applications Data Flow. The flow of applications data is primarily through commercially sponsored channels such as the technical trade press, catalogs, exhibitions, advertising, product manuals, vendor data services, and training programs and materials. Table III-B-1 and Figure III-B-4 illustrate the training data flow process. Product manufacturers generate training materials containing the minimum data required for product use. Products may be capital or consumer goods, subsystems, or large systems. The data used may come from one or all three of the following source types: (1) standard reference material, (2) formally organized internal product data systems, and (3) training material produced by commercial firms (such as Howard Sams & Co.; McGraw-Hill, Inc.; and Industrial Education Institute). Some manufacturers tend to primarily use internally developed materials and other externally available materials, but few use both. For example, Ford Motor uses mostly outside sources, while GM and Chrysler use internal sources and make a business out of training through training institute operations. Most large organizations use central training officers (staff executives) to coordinate organization-wide training activities, and to assure efficient use of data resources for training material development. Federal agencies fail to do this to any extent, although Project Teach (listed in Table III-B-1) is one effort bent in this direction.

The findings shown in this table resulted from a specific survey/ analysis made through structured interviews conducted at the annual meeting of the American Society of Training Directors (now called the American Society for Training and Development).

Table III-B-1 Training Data Development

Training Materials Used	Organization	Training Materials Developed
published training materials instrument and equipment manuals duPont Industrial Library	Organic Chemicals Division Olin-Mathieson Corporation	None
handbooks, trade journals, product manuals, corporate technical data systems	Sylvania Electric Products, Inc. Electronic Systems Division	equipment and system manuals for customers
National Agriculture Library Forest Service Facilities	Forest Service U.S. Department of Agriculture	training manuals for Job Corps forestry operations
DoD training manuals, including programmed instruction courses	Project Teach (Training & Education Clearinghouse) U.S. Air Force	catalog of DoD training materials being developed
NSIA members' training materials	Project Aristotle (Annual Review of Information Sources on Training, Learning and Education) National Security Industry Association	annual review of state of availability of training materials for defense contractors' use
technical publications, engineering documents, specifications, research reports, government reports, equipment-test data	IBM Corporation	system manuals, employee training programs
primarily commercially published training materials; except for data produced internally for automotive and Philco system manuals	Ford Motor Company	employee training materials automotive and Philco system manuals
technical publications, engineering documents, specifications, research reports, government reports, equipment-test data	General Motors Corporation	materials for General Motors Institute and manuals for produced system
primarily internal documents, libraries, and staff.	Standard Oil Company of New Jersey	internal training program materials and systems for instructing techniques in exploration, production, transportation, refining, marketing, management, tanker ship operations, etc.
technical manuals, vendor data, subsystem training manuals, engineering reports, drawings	Link-Tempo-Volt Corporation	missile system manuals

Figure III-B-4 Training Data Flow

C Representative Problems

The problems identified in the course of this survey/analysis of the ten selected fields of science and technology fell into three groups--those associated with the management of basic, of developmental data, and of applications data

- Basic Data The problems in this area fall into two subcategories associated with the requirements for (1) current awareness of new data which are of potential utility in research or other discipline-oriented activity, and (2) retrospective search for data resulting from previous research. Satisfying the first requirement is a problem not only because of the continual increase in the amount of research performed, but also because of the current interweaving of disciplines within the physical, life, earth, and social sciences. Satisfying the second requirement entails meeting the needs of two groups--research-discipline activities which need data for application in new investigations, and mission-developmental data activities. One problem encountered in retrospective data search and retrieval is the tedious two-step process involved, i.e., a publication search followed by a data search within the publications. A second hindrance is the three- to five-year time lag between data generation and publication. The two-step process for a data search could be shortened considerably by requiring the indexing of data contained in basic journals within all scientific and technical disciplines.
- Developmental Data A serious problem in this area is the inability of scientists and technologists engaged in mission-developmental data activities to obtain the degree of data accuracy and precision required for accomplishment of their tasks More data centers are needed to provide adequately evaluated and qualified data which are of broad developmental utility. A study of the coverage, quality, and quantity of data required by developmental data activities should be conducted Until adequate data centers are designed, each project, mission or task must meet its own data requirements and must provide data support services that are sufficiently dynamic to keep pace with changing requirements
- Applications Data Two primary problems were identified within applications data activities. (1) duplication of data dissemination effort which is partly due to competition; and (2) inadequate data flow for meeting the requirements of all user communities The first problem cited probably can only be solved within those Federal and Federally-sponsored programs which generate applications data. In such setting the coordination and identification of data commonality which are required to alleviate the problem can be achieved relatively easily The second problem may be solved as the number of commercial vendor data companies such as Information Systems

Division of McGraw-Hill, Inc., and the Information Handling Services, Inc., increases. The growth of such companies, which make a profit from selling or promoting commercial vendor data, will increase availability of and control over disseminated data.

D. The Future Outlook

It is not within the scope of this section to present all of the significant developments which will affect the future outlook for scientific and technical data management. Volume I of this report handles this question in the framework of the time-phased plan. This section, however, does present a brief assessment of future possibilities, based upon the findings of the ten write-ups.

- Basic Data. Programs dealing with the problem of indexing data content of literature will evolve from present and future studies of data management, in support of research. It was forecasted by some observers that eventually researchers will seek modes of data dissemination outside the traditional publication mode, and will develop data centers for specialized fields which would facilitate automation of experiments, as well as interactive data retrieval, storage, and computation functions.
- Developmental Data. In accordance with the trend toward developing models of phenomena and systems, more developmental data of broad utility will be centrally used as the basis for micromodeling. This will facilitate more data prediction and less generation of data which is useable for only one purpose. A closer relationship will be established between directed research activities and early-developmental activities, as they tend to use the same data resources more and more.
- Applications Data. The increase in number and sophistication of commercial vendor data services portends an increase in on-site terminals for remote retrieval of vendor data in support of discipline-research, mission-developmental, and applications-product activities. Eventually vendor data will be used in computer-aided and computer design and modeling of systems.

PART B

I. STRUCTURE AND CONTENT

The conduct of a large-scale pioneering census, by necessity, involves formulation and application of broad structuring concepts. For example, in Parts A and C of this volume, the concepts of communities of scientific and technical interest and of formal data efforts are used as key structuring concepts. These approaches proved adequate to obtain the broad-spectrum perspective required to meet the major objectives of the census effort. However, it is also desirable to assure that the defoliations and generalizations required to develop these structuring concepts and apply them in a census effort do not mask out significant attributes of data activities. Consequently, a series of survey probes was conducted of selected areas of scientific and technical data activities. In addition to assembling a limited amount of census-like information, these probes also generated information relevant to specific issues or problems which are discussed in Volume I.

Specific probes and surveys reported in this section are as follows:

- Survey of Centrally Coordinated Data Activities of Medical Schools and Related Research Institutions--A Probe of Practices, Trends, and Problems in Data Handling in a Specific Type of Research Institution

- Survey of Data Activities of Selected Professional Societies and Trade Associations--A Selected Probe of Institutional Capabilities and Plans

- Survey of Commercial Data Centers Which Process Scientific and Technical Data--A Selected Probe of the Practices, Trends, and Problems in a Selected Type of Data Processing Organization

- Summary of Scientific and Technical Data Files of the Department of the Army--A Selected Inventorying Probe of Existing Data Resources

- Review of Current Equipment Capabilities for Scientific and Technical Data Handling

As a set, these probes begin to illuminate the complexity and state of flux currently found in scientific and technical data activities. In summary, they indicate that existing organizational structures and human competencies are over-extended in terms of their abilities to effectively accommodate evolving data management and data handling needs. In addition, alleviation of this situation does not appear imminent, for only now are the individuals and organizations affected beginning to recognize the gravity of the situation. To date, attempts to improve data management and data handling operations have been rendered largely ineffectual due to crisis-type actions and piecemeal approaches to broad problems. Significant improvements cannot be expected until needs are better defined, organizational responsibilities are identified and accepted, and increased funds are made available to support the effort required to alleviate existing problems.

Science Communication
Washington, D. C. 200 07

II. SURVEY OF CENTRALLY COORDINATED DATA ACTIVITIES OF MEDICAL SCHOOLS AND RELATED RESEARCH INSTITUTIONS

-- A Probe of Practices, Trends, and Problems in Data Handling in a Specific Type of Research Institution

A. Statement of Purpose

A survey was undertaken to determine the present status of centrally coordinated medical and scientific data activities within medical research schools and related institutions. The survey was also intended to probe the broader question of current problems involving medical data. This probe supplements the cross-sectional study conducted of data activities in the various areas of science and technology. Specifically, it has supported preparation of a current status write-up of data activities in the biomedical sciences.

A specialized questionnaire was sent to the ninety-five medical schools operating or under development in the United States and Puerto Rico, as well as to nine other medical institutions such as major hospitals and research foundations. The nine medical institutions were added to the sample because the research functions within selected hospitals, research foundations, etc. are quite similar to those in medical schools. The survey yielded meaningful compilations of the current services and processing capacities which are available in these institutions. In addition, it has unearthed many serious problems facing medical data operations as well as forecasts of anticipated trends and changes within teaching institutions, hospitals, and private practice.

B. Survey Approach and Response

Of the 104 institutions which received questionnaires, fifty-three (50.9%) have responded. Of the fifty-one not returning questionnaires, ten wrote letters explaining that their data services have

only recently begun and that, therefore, their experience is too limited to answer the questions. In addition, two (VA Hospital, D. C., and the University of Arizona) explained that at present they are not conducting any of the data activities described. Three other schools stated that there are only two-year programs available presently within their institutions.

C. Summary of Findings

1. Description of Data Activities

Many medical schools presently have no internal computing facilities, but have access to the main university computing center. Several reported having a small internal equipment capability, but using the main university computers for bulk computing. On the other hand, a few schools like St. Louis University Medical School have extensive internal facilities which they share with local hospitals.

The most common data service offered by these schools is a computing service. Other commonly available services are data analysis, data reduction, experiment design, computerized data acquisition, systems analysis, data centers, systems programming and maintenance, data storage and retrieval program libraries, mathematical consultation, on-line plotting, "quickie" courses in programming languages, and data archive operations.

For purposes of general classification, the data activities performed by medical research institutions have been divided into four categories--laboratory data reduction and analysis, epidemiological correlations, clinical data coding and processing, and hospital patient data processing. Table II-C-1 gives a breakdown of responding centers in terms of the data activities conducted.

The percentage of users listed as external to the responding institutions varies greatly--from 1% to 100%. Most of those respondents reporting a 100% external user population, however, are not standard medical schools, but hospitals and research foundations. The median percentage of external users reported is 3%.

TABLE II-C-1

DATA ACTIVITIES CONDUCTED

Activity	No. of Centers	% of Total Respondents *
Laboratory Data Reduction & Analysis	38	71.7%
Epidemiological Correlations	29	54.7%
Clinical Data Coding & Processing	33	62.2%
Hospital Patient Data Processing	21	39.6%

SPECIFIC SERVICES OFFERED

Service	No. of Centers	% of Total Respondents *
Experiment Design	32	60.3%
Computerized Data Acquisition	3	5.6%
Data Analysis	24	45.2%
Computing Services	37	69.8%
Data Storage and Retrieval	5	9.4%

EXTENT OF AUTOMATION

Extent	No. of Centers	% of Total Respondents **
Total	6	11.3%
Substantial	15	28.3%
Limited	19	35.8%
None	3	5.6%

* The percentages shown total more than 100% because many centers perform more than one of the services listed.
** The remaining 19% did not respond to the question.

The specific services offered to users by medical computing centers have been divided into five classifications--experiment design, computerized data acquisition, data analysis, computing services, and data storage and retrieval. The statistics compiled from the responding centers on the services they offer are shown in Table II-C-1. Also exhibited in Table II-C-1 are the results of our inquiry concerning the extent of automation currently existing in the centers.

It was requested that the centers provide estimates of the number of professional and support personnel employed. The range of professional staff members reported is from one to seventy-five, while the number of support staff members varies from one to one hundred. The average (mean) number of professional staff members in the responding institutions is nine, while the average support staff has 13 members.

The computing centers were also asked to supply three indices of current annual growth rate--number of users serviced, size of staff, and amount of data handled. The returns are somewhat ambiguous, since many have only recently started services, and many others failed to itemize which index of growth they were reporting. The percent of growth in terms of users serviced ranges from 0% to 200%, with a median growth of 35%. The growth rate in the size of staff varies from 0% to 200% (median of 50%), while the growth in terms of data handled varies from 0% to 300% (median of 50%). These replies reveal dramatically the boom taking place at present in medical data activities.

The survey reveals that centrally coordinated data activities in medical research institutions are quite numerous. It was found that virtually every medical research institution utilizes a data processing center. Schools which are currently under development usually have plans to establish centrally coordinated data activities in the future. In addition, many of the respondents demonstrate a comprehensive knowledge of the potentials of computer equipment and systems. Ambitious planning, however, is often frustrated by the problems discussed in the following sections of this report. Thus, the willingness and potential to innovate seem to exist, but the means to innovate often do not.

A few exceptions to the preceding generalization are noted within certain older, well-endowed medical institutions. Outstanding instances of modern, well-coordinated biomedical data activities exist at Harvard University, Baylor University, the University of California at Los Angeles, the University of Maryland, and the University of Texas Southwestern Medical School. Ten other responding institutions report having large, well-equipped facilities, and many others are presently developing highly automated centers.

2. Problems Confronting Medical Data Activities

There is great diversity of opinion as to which problems facing medical data operations are most serious. Those problems mentioned by the respondents have been divided into several main groups according to topic.

(a) Personnel Problems. A great majority of the problems mentioned concern personnel. Most of the current personnel problems could be ultimately traced to the gap, caused by advancing technology, between equipment capabilities and personnel capacity to make optimal use of such equipment. Most often cited is the need for training of computer-oriented paramedical personnel. It seems that the present structure of most institutions fails to meet modern needs. Typical hospital personnel need to be re-educated for interface with the new services. Data processing specialists usually have no conception of biological problems, and bioscientists are often unwilling to learn computation. Therefore, training programs are desperately needed to produce personnel conversant with both computing and medical disciplines.

Some personnel problems are attitudinal. For example, systems analysts often underestimate the serious nature of medical problems, and their optimism can be detrimental to progress. Medical information scientists often demonstrate tunnel vision in failing to recognize that problems they consider unique to the health sciences might be partially solved by exploring other fields.

Partial solutions offered to these problems are as follows:

- Massive training programs in computer science and mathematical modeling in the life sciences at the Bachelor's level.
- Educational programs at the university graduate level designed to produce medical documentalists and information processors.
- Development by biomedical institutions of an in-house competence in the computational sciences.
- Establishment of a training program for computer staffs whereby the staff can spend more time on individual researchers' problems, and attend conferences within the medical school. One staff member should concentrate on automating each department's data procedures and on developing better updating procedures.

(b) Financial Problems. Financial problems rank second in frequency. The factors most often mentioned are the high cost of equipment, of computer services, and of space to accommodate central computing facilities. The difficulty of providing adequate speed of response at a low cost, with protection against machine failure, was cited as a serious multi-faceted problem to be corrected. An additional concern is the high cost of the introduction of new methods. Since normal operations cannot stop while new ones are being initiated, often there is temporarily total duplication of effort. Compounding these problems is the current administrative squeeze on funds in many institutions.

There is much disagreement among respondents about the ideal solutions to these financial problems. Many feel that more Federal support is needed. One respondent states that such Federal support should be only for non-fee-for-service operations. Another says that grants should be made available to underdeveloped institutions, rather than to thriving institutions. Still another participant, however, feels that increased Federal funding is not the answer. His recommendations are a decrease in the cost of computers and an increase in state funds for educational computing.

(c) Equipment and Software Problems. Equipment and software difficulties were mentioned quite often by respondents. Many of the problems can be traced to financial difficulties, or are local within certain institutions. Several, however, are of national scale and importance. The unreliability of computer equipment for continuous on-line operations presents problems, because it necessitates back-up systems for crucial programs. A software problem cited is the unavailability of simple interpretative programs that can be easily learned by users. The large variety of analog data collected by researchers within an institution complicates the service of individual needs, since transforming such a variety of analog data into digital data is quite difficult. In addition, technological advances are needed to permit a remote user, at his discretion, to engage the computer in batch or computing modes, or a conversational interaction mode. Compounding all hardware problems are the lack of standardization among vendors of computing devices, and the differences in programming specifications from one machine to the next.

Ready answers to these problems are scarce. The only recommendation offered for any of the equipment and software problems was that hardware developers should concentrate on developing peripheral terminals and small peripheral processors which can be easily linked into a functional net. Such equipment would greatly facilitate future development of systems.

(d) Coordination Problems. Many difficulties stem from lack of coordination within existing systems. It would be highly desirable, for example, to share operational information systems among biomedical computer groups. Unfortunately, however, the efforts are seldom coordinated, even within single institutions. Duplication of effort and duplication of errors occur quite frequently. If a sharing system were established, centers could communicate about programs, programming techniques, and comparative evaluations of computer hardware--thus sharing the benefits of experience and avoiding many of the errors.

In order to accomplish such coordination, leadership in national programs should be by experts in both computing and medical problems. One or more high-caliber periodicals could be established which are

devoted to biomedical computing (e. g., "Computers and Biomedical Research"). In addition, active avenues of communication must be established among isolated data programs in each university. Perhaps a university-sponsored bulletin on data programs would accomplish this.

(e) Administrative Problems. Lack of cooperation from university administrations often thwarts progress in data activities. Sometimes the appraisal by administrators of the role and cost of data processing in medicine is incomplete and ill-informed. Many top-level medical personnel are skeptical because the individual variability of data and the complexity of services provided in medical institutions do not readily yield to simplified modeling.

To solve these administrative problems, it is suggested that administrative officials be provided a series of realistic order-of-magnitude figures on the budget needed to implement the following: (1) small-scale data handling projects, (2) a central computer with a few external terminals for limited operation, and (3) large-scale multi-purpose time-sharing enterprises. They should also be given the amount of time required to get such systems into useful operation. The use of modern management techniques within the health care and health science fields should be encouraged, since it would greatly facilitate the transition.

(f) Input Data Problems. The quality of input data is also regarded as a critical problem in medical data activities. Serious errors often occur when data are entered in source documents, which are later converted into tab cards. The errors are made by those collecting the data, including professionals, and they occur regardless of the type of source document. The mistakes are usually of greater magnitude when data must be coded before submitting them to analysis. Errors also occur when data are entered from a remote terminal to a computer. To solve this problem, computer routines for error detection should be established so that data which do not fulfill certain criteria of acceptability are rejected and returned to the user for correction.

An additional problem was mentioned by several respondents which does not fall under any of the preceding categories. There seems to be a current overemphasis on demonstration projects for development of hardware and bigger systems. There should be increased funding for applications projects instead, since knowledge of applications is vital for the medical field during this stage.

3. Opinions Concerning Key Issues

Previous discussions with medical data personnel during project workshops have generated several important issues. Two of these were chosen for inclusion in the questionnaire for this survey. One concerns standard medical nomenclature, and the other concerns data requirements of medical workers.

(a) Standard Medical Nomenclature Issue. We first asked participants to state an opinion about the desirability of establishing a standard medical nomenclature and coding system. We then asked that those answering affirmatively suggest the name of a pre-existing group, or the composition of a newly formed group, to set up and be responsible for such a system.

Thirty-four (64.1%) of the respondents feel that a standard nomenclature and coding system is needed for use in data systems. Nine (16.9%) say there is no such need, and eight did not respond. One of the respondents feels that standardization is not presently feasible due to the primitive state of data collection. Another says that standardization has caused considerable error in the analysis of clinical data. Still another opinion offered is that there will be standardization of procedures and systems, but not of codes and nomenclature, for "it has been proven over and over again that standardization via codes and nomenclature (is) destined to be inadequate..." Another dissenter feels that "standard should read recommended. A total system of nomenclature should be the goal, including automatic methods of updating files to reflect continuing revisions in recommended nomenclature."

Among the respondents which answered affirmatively, there was great diversity of opinion about what group should be responsible for such a system. Virtually no consensus can be detected among

respondents concerning the ideal composition of such a group. Most respondents did agree in general, however, that the group should be interdisciplinary, and that it should combine both medical and computer personnel.

(b) Medical Data Requirements Issue. This question has been central during our study: "Are the data requirements of workers in this field adequately defined to permit the development of improved data services?". When medical workers were queried concerning this, nine respondents said "yes," thirty-one said "no," and the remaining thirteen did not answer. The response shows that this is an area where further study should be conducted.

Suggested research projects which should be undertaken to identify these data requirements in medicine are quite varied. The following are some of the more significant suggestions:

- A series of interdisciplinary meetings should be held which include physicians, hospital administrators, medical record librarians, and computer scientists. These meetings should be organized to identify requirements in specific areas (e.g., medical records, selection of treatments, scheduling of tasks, simulation studies, etc.).
- An attempt must be made to interview in-depth leading practitioners in the medical field in order to determine their specific requirements. A day-by-day, on-the-job analysis of data and information requirements may be necessary.
- Problems to be defined include the need for computer-oriented files of research data of individual workers, available to the scientists through terminals in their own laboratories. Other areas of interest would be the utility of special artificial languages designed to serve the needs of biological workers.

- More knowledge is needed about the relative amount of information which should be collected in fixed-field, coded, or standard nomenclature rather than in variable length free text without standardization.
- Research might be undertaken to determine the relative need of types of services in medicine. This study could be done on selected subpopulations.
- A great deal of work must be done on gathering data about "normal" individuals in many categories (age, sex, body build), obtaining follow-up information at periodic intervals, and using this body of information for test purposes.
- Much data is descriptive (narrative) and does not lend itself easily to coding for a computer. Research is needed in methods of training physicians to modify their language to make it more acceptable for coding. In addition, a study should be undertaken about enabling the computer to better accept the present language of the physician.
- Research will have to proceed from the particular to the general. No clear evidence exists that these data requirements are homogeneous; therefore, detailed studies in each discipline seem necessary before general conclusions can be drawn.
- In spite of legal requirements, the definition of a patient's medical record apparently is not clear. It is likely that less research and more development are called for. Information scientists should work in parallel with health personnel to define the requirements.
- Studies are needed about what information doctors really use in the decision-making process. One goal should be the elimination of ambiguity and redundancy in medical reports. A careful analysis should be made of the value and use of data contained in the present standard medical record.
- Because of the ever changing and expanding nature of data gathering, there will never be a complete definition of the data requirements of workers in medicine. Therefore, the

research undertaken to identify these requirements must be a continuous program with periodic conventions and publications for the determination, solution, and dissemination of the requirements. Such a program would have to be national in scope and broadly composed of members from all phases of medical research.

- Evaluation and utilization of published data involve a complex process, often dealing with such intangibles as "reputation" of the authors, the journal, or its editors. The traditional "scholarly paper" method of reporting has merit in terms of these factors, which would be lost in any mechanical process of condensing and compiling data from multiple sources. Such a compilation could be of enormous value, if it ever becomes possible to exclude or allow for the "source" factors.

- If each author who contributes to a medical publication were required to provide a summary of the paper, including a summary of pertinent data, and if the editor could be paid for evaluation and condensation of this abstract, then one might have readily available "data" abstracts for storage and retrieval.

- A coding and abstracting system might be established on a national scale. The editorial board of that abstracting and coding service would invite certain journals, selected according to quite restrictive scientific standards, to adapt themselves in format to the needs of the abstracting service, while others would not.

4. Forecasted Trends and Changes

Participants were asked to predict trends and changes in data handling within the next ten years in teaching institutions, hospitals, and private practice.

(a) Changes in Teaching Institutions. The following innovations are foreseen within teaching institutions in the next ten years:

- Greater use of on-line terminals for immediate analysis of data generated by the student laboratories of physiology, biochemistry, pharmacology, etc. Hybrid computer systems may be suitable.
- Use of computer techniques to simulate specific clinical situations and to evaluate the medical students' performance during these simulated experiences.
- Extensive use of computers for testing mathematical models of biological systems and for control of experiments.
- Courses in programming, mathematics, and the elements of analog computers, and basic techniques of information sciences.
- Development of large-scale medical information systems.
- Broader acceptance and use of television and computer-programmed instruction will allow for increased enrollments, more standardized instruction among institutions, and greater personal attention by instructors.

(b) Changes in Hospitals. Within hospitals, the following trends and changes are predicted:

- Establishment of a computer-based data communication network.
- Use of computers for screening medical histories provided by patients.
- Use of computers for automatic scheduling of tasks, for allocation of personnel and equipment, for billing, and for inventory control.
- Use of computers to assist the physician in diagnosis and in selection of treatments and drugs.
- Extensive physiological monitoring with on-line computer processing of data.
- Use of computers for automatic analysis of laboratory samples and for automatic interpretation and reporting of results.

- Continuous compilation of patient data for epidemiological predictions.
- Increased use of non-M. D. health care personnel for obtaining patient historical and laboratory data for input to computer memory. Thus, non-M. D. personnel will gradually take over more and more of the diagnostic function, leaving the physician the therapeutic function.
- Large files available for terminal interrogation concerning blood donors, poison centers, cytology, etc.

(c) Changes in Private Practice. The following effects are predicted in private practice:

- More data on hospitalized patients will be available to physicians on an automatic basis. Use of terminals in doctors' offices which are hooked up with hospital computer facilities will be common. Telephone terminals will be in limited use.
- Use of central facilities for fiscal operations.
- Central registry of patients in local community, including current diagnosis and medications (medical record data banks).

The respondents seemed to reach a general consensus that innovations within the next ten years will occur first and most extensively within teaching institutions, next and less extensively at hospitals, and last and least extensively at the private practice level.

III. SURVEY OF DATA ACTIVITIES OF SELECTED PROFESSIONAL SOCIETIES AND TRADE ASSOCIATIONS

-- A Selected Probe of Institutional Capabilities and Plans

A. Statement of Purpose

A survey was conducted among selected professional societies and trade associations in order to ascertain the current role played by such organizations with regard to scientific and technical data. It seems possible that these pre-existing organizations of data users may eventually become key links within data systems of national significance. Since these organizations are largely discipline- or industry-oriented, they might serve as convenient channels for communication with developers or users of future data systems who have common interests and data needs.

B. Survey Approach and Response

The sample was chosen from a population of 3500 national associations presented within the 1967 edition of the Directory of National Trade and Professional Associations of the United States. Our final sample of 171 organizations resulted from the selection of those groups within the population which met certain size criteria in their staffs and memberships, and which were concerned with scientific and technical matters. Specifically, all professional societies having staffs of more than five and memberships of more than 5, 000 were included. In addition, all trade associations with staffs of more than five and memberships over 50 were selected for the sample. The candidates chosen by this process were then screened to assure that our final sample represented the scientific and technical scope included in our project. The resultant group consists of 121 professional societies and 50 trade associations.

Thus, due to the sampling methods used, our survey results constitute an assessment of the current status and possible future capabilities for data handling within the nation's largest trade and professional associations with an interest in science and technology, as well as technology-based industries. Since an organization's size is usually an index of its financial backing and staff capability, and such groups often are more progressive in terms of the initiation of new programs, we feel that the organizations in our sample constitute likely candidates for implementing future data systems of national importance.

Of the 171 organizations receiving questionnaires, sixty-four (37.4%) returned usable responses. Eighteen trade associations and forty-six professional societies comprise the responding sample. Of the 107 remaining groups which did not reply satisfactorily, eighty-three sent no answer at all, and twenty-four sent letters or postcards explaining that the questions did not seem applicable to their activities.

C. Summary of Findings

1. Unanticipated Factors Affecting Results

Many different levels of comprehension of the subject of our survey were displayed by the responding organizations. A few requested further explication of the term "scientific and technical data". Since we provided no precise definition of the term in our questionnaire, diverse interpretations were reflected in the survey results. Most officials of professional societies and trade associations do not seem to differentiate between "data" and "information" in their thinking; nor do they readily sense the difference between "data handling" and "document handling". One might conclude from the negative results obtained by a complex questionnaire that future correspondence with such groups should be educational and only progressively complex.

On the other hand, a second factor affecting results was the feeling among organization officials that the content of the questionnaire was of no official concern to the society or association. Several respondents indicated that although officials of their organizations have

extensive knowledge of the data interests and activities of the members, the society "as a society" has no responsibility in these areas ot its members. Apparently the administrations of many societies and associations have not yet deemed it necessary or desirable for the organization, as such. to play a significant role in the collection, storage, retrieval, and dissemination of data for its members, except through the conventional means of journal publication and conduct of meetings.

2. Current Data Activities

(a) Data Services. Only thirty-eight (59.3%) of the respondents conduct data activities, other than publishing journals or bulletins containing data. Many of the respondents conduct activities which fall into more than one category of data activity. A breakdown of the reported activities is presented in Table III-C-1. The publication of journals containing data was not included in these totals because such activities usually do not entail handling data as such, but merely involve reviewing and publishing contributed articles as units It should be noted that a great majority of the respondents (85.9%) reported publishing journals or trade bulletins. In fact, some cited the journal publishing activity as the primary motive for the organization's existence, and as its central source of funds.

(b) Groups and Committees. The concern of an organization with a particular problem or subject is often indicated by its establishment of a committee to deal with the matter. We asked the respondents to provide the names of any groups within their organizations whose pri-ıary concern is scientific and technical data or the transfer of data. Thirty (46.9%) of the participants reported having a committee on scientific and technical data. The following committees were mentioned:

SOCIETIES

- Tec' ical Council, Air Pollution Control Association
- Data Processing Coordinator, American Association of Medical Record Librarians

TABLE III-C-1

DATA ACTIVITIES CONDUCTED BY RESPONDENTS

Activity	No. of Organizations	% of * Total Respondents	Professional Societies	Trade Associations
Publishing Data Sheets, Technical Manuals, Handbooks	19	29.7%	13	6
Generating or Publishing Specifications or Standards	18	28.1%	5	13
Generating Data Through Research	30	46.9%	16	14
Determining Data Needs of Members	14	21.9%	10	4
Operating a Technical Data Center	2	3.1%	0	2

* The percentages shown total more than 100% because many respondents conduct more than one of the activities listed.

- Committee on Analytical Reagents, American Chemical Society
- Department of Research and Statistics, American Chiropractic Association
- Survey Research Department, American Medical Association
- Committee on Standards, American Optometric Association
- Committee on Drugs Standards, Analysis and Control; American Pharmaceutical Association
- Information Processing Project, American Psychiatric Association
- Project on Scientific Information Exchange in Psychology, American Psychological Association
- Evaluation and Standards Committee, American Public Health Association
- Standards Committee; American Society of Heating, Refrigerating, and Air-conditioning Engineers
- Committee on Drug Information Services, American Society of Hospital Pharmacists
- Standards Committee, American Society of Tool and Manufacturing Engineers
- Handbook Committee, Illuminating Engineering Society
- Data Series Committee, American Institute of Planners

ASSOCIATIONS

- National Aerospace Standards Committee, Aerospace Industries Association of America
- Technical Research Committee, Alloy Casting Institute
- Technical Division, Aluminum Association
- Subcommittee on Technical Data, Division of Refining, American Petroleum Institute

- Technical Services Division, American Plywood Association
- Handbook Committee, American Society for Metals
- Technical Activities Committee, American Welding Society
- Technical Information Services Committee, Federation of Societies for Paint Technology
- Standards Committee, In-Plant Powder Metallurgy Association
- Standards Committee, Metal Powder Industries Federation
- Codes and Standards Committee, National Electrical Manufacturers Association
- Quality Control Section, Pharmaceutical Manufacturers Association
- All committees, U.S. of America Standards Institute
- Technical Committee, Western Wood Products Association
- Standards Advisory Committee, Technical Service Advisory Committee; Copper Development Association

These committees could serve as foundations for the development of future data activities. Only five organizations reported the establishment of a committee on the transfer of data, however. These replies seem to demonstrate a lack of emphasis on these matters within over half of the organizations.

(c) Meetings and Symposia. Additional evidence which supports the preceding conclusion concerning lack of current emphasis is the dearth of meetings and symposia on data management held by these organizations within the last three years. Only thirty-two groups reported holding such meetings, and most of the reported meetings were only secondarily concerned with the subject of data management, data services, and data systems.

3. Current Issues and Problems

Participants were requested to identify scientific and technical data needs which are not satisfied by current or planned data

services. The following individual data needs were mentioned:

- Much more research on urban development models, transportation demand, required levels of social services, and techniques for evaluation of urban development patterns needs to be performed in order for us to refine and become strategic about our data requirements. (American Institute of Planners)
- There is an urgent need for more thermodynamic data on hydrocarbons and their related compounds. (American Petroleum Institute)
- Additional data are needed on the characteristics of certain aluminum alloys and products. (The Aluminum Association)
- A Handbook for Optometry (on visual science) analogous to the Illuminating Engineers Handbook is needed. (American Optometric Association)
- Additional data are needed from clinical trials to determine the efficiency of various therapeutic methods used on specific types of cases and for prediction and prevention studies. (American Chiropractic Association)
- Facilities are needed for identifying psychiatrists having particular subject-area competence. (American Psychiatric Association)
- Information is needed on past and current research projects and their findings in disciplines related to forest products, forest economics and forestry. (Western Wood Products Association)

Other replies contained suggestions for general improvement in the current capabilities for data acquisition. Significant recommendations include the following:

- There is a need for better standard terminology, definitions, and nomenclature in technical abstracting and selective dissemination of information programs.

- All engineering periodicals should perform source indexing.

- Specific referral centers which are linked to the National Referral Center for Science and Technology should be set up in order to provide information about the locations of documents. The centers should be discipline-oriented and should be able to tell users which libraries have copies of desired documents.

- A central clearinghouse for military and industry standards and specifications documents, similar to Defense Documentation Center for technical reports, should be established.

- The retrieval system of the National Library of Medicine needs to be codified in order to make resources there of maximum use to public health specialists.

- The urgent need is the tran·lation and implementation o' technology to operations. Presently executed data programs are beyond immediate needs, and have been formed to instigate technical growth within the industry, assuming industry can be persuaded to use the technology.

Many problems encountered by members of the responding organizations in their use of the national scientific and technical data resource coincide with those voiced by participants in other phases of our study. Those problems of the greatest consequence and the suggested ʾans for solving them are as follows:

- The information explosion has made comprehensive coverage of needed information almost impossible. The rate at which technical data are being generated has outmoded the classical system for handling data. Much published material is repetitive and presents no new data. A partial solution to these problems lies in more evaluation, selection, and summarization of substantive literature.

- One problem is to identify and retrieve only that information which is pertinent to the matter at hand; the other is to identify the material which has true merit. These

difficult qualitative judgments cannot be entirely delegated to a central data screening agent, but steps in that direction seem necessary in order to cope with the ever-increasing data resource. An abstracting service cutting across all scientific and technical disciplines and able to respond to highly specific, programmed inquiries could be the solution at the national level.

- There is insufficient awareness of the information facilities that are available, where they are located, what they contain, and how to make use of them. Promotional and educational programs about information resources should be initiated to correct this situation.
- There is a lack of communication among organizations and/or agencies about what information is at hand and what information is needed. Some organizations are unwilling to share the results of studies, surveys and other data collections. Regular channels of communication should be established in order to pool information.
- Strictly observed, the copyright laws are an impediment to the best use of our data resource. The users of abstracts have difficulty obtaining complete texts of articles containing scientific and technical data which are cited in them, due to copyright restrictions.
- On the other hand, the pending revision of the copyright law seems to give little or no protection to owners of copyrights against promiscuous "fair use" copying or against integration of data into electronic storage. Ways must be sought to stop publishing information in forms which can be pirated without reference to the copyright.
- Methods are needed whereby the cost of dissemination and processing of data can be reduced. The advent of a competitive cold type-setting process appears to offer an opportunity to reduce such costs. The technique provides, as a by-product of the printing process, full-text machine readable data at little or no extra cost.

- Voluntary international commercial standards should be promoted, and a clearinghouse for commercial and procurement standards should be established. The passage of legislation currently being considered would accomplish these objectives, and would aid the national technical purpose. In addition, legislation to study the advantages and disadvantages of the metric system should be supported.
- There is often underemphasis on information-gathering as an integral part of professional role behavior. To correct this situation, emphasis on problem solving, as contrasted to information mastery, should be increased during professional training.

Several organizations mentioned individual problems within their disciplines. A few of the noteworthy ones follow:

- Establishment of additional medical library facilities -- smaller versions of NLM across the country--should be considered. (Association of Military Surgeons of the United States)
- More unification of the data systems used by state and local health departments is needed in order to accumulate comparable data from these institutions. In addition, a single birth to death health record for every individual would provide a much improved base for accumulating health data. (American Public Health Association)
- It would be helpful if there were a central source of engineering data. (American Petroleum Institute)
- Our first requirement is the development of a comprehensive index of standards. This must include approved national standards, technical, professional, and trade association standards, and proposed standards in the developmental stage. Such an index would form the basis for an information retrieval system. (U. S. of America Standards Institute)

- The lack of standard terminology thwarts research efforts. A proposed standard terminology should be developed under the aegis of an organization like the National Academy of Sciences. (Society for Industrial and Applied Mathematics)

4. Planned Activities

A few (14%) of the responding organizations have plans to improve or initiate data activities of various types in the near future. The following plans were mentioned:

- Installation of remote terminals to the central computer, thus enabling on-line retrieval of information (Copper Development Association)
- Implementation of a new information storage and retrieval system (U.S. of America Standards Institute)
- Establishing facilities for accumulating state-of-the-art data from worldwide sources and disseminating the data through subscription services (American Foundrymen's Association)
- Acquiring capital equipment for use in a computerized information storage and retrieval system; and making a study to determine the technical information requirements of the members today and ten years hence (American Society of Tool and Manufacturing Engineers)
- Extending the scope of the standards now published (Dairy and Food Industries Supply Association)
- Broadening the areas covered in research projects (American Chiropractic Association)
- Creating a computer processable Drug Products Information File (American Society of Hospital Pharmacists)
- Conducting research on future computer typesetting of publications and possible selective dissemination of literature (American Chemical Society)

5. Suggested Means for Handling Medical Data

Participants were requested to suggest improved methods for the selection, evaluation, condensation, and dissemination of medical data. Some of the significant recommendations are as follows:

- The NASA SID system should be adapted for the medical field. An asset of the system consists of the feedback of information provided by the scientist in evaluating the usefulness of abstracted reports.
- Many new journals are presently being published which are constantly in search of articles without concern for quality and novelty of the content. This trend must be reversed.
- It would be desirable to decrease the time lag between submission of papers and their publication.
- Perhaps there could be an accreditation body for journals. Requirements for accreditation could include these:

  (a) An editor for literal composition, grammar and style.
  (b) An editor for numerical composition and style.
- The publication of data in a tabular form with a minimum of interpretation by the investigator and a statement about the methods used would prevent much duplication and facilitate use.
- Computer-oriented bibliographical and abstracting services will become more useful when combined with a means for facsimile transmission of source documents. It appears that computer-based indices and retrieval systems will only be a partial solution to document proliferation if libraries in every geographical area must accumulate all of the medical literature whether there is a local demand or not. Indexing which evolves from the word usage not only of authors but also of readers of articles will be needed.
- Various approaches must be evaluated. In addition to the schema for MEDLARS, others should be pursued, e.g., SDI, natural language retrieval, conversational mode processing, computer-assisted instruction methods.

Science Communication
Washington, D. C. 200 07
COSATI Data Activities Study
Final Report - F44620-67-C-0022 30 April 1968

IV. SURVEY OF COMMERCIAL DATA CENTERS WHICH PROCESS SCIENTIFIC AND TECHNICAL DATA -- A Selected Probe of the Practices, Trends, and Problems in a Selected Type of Data Processing Organization

A. Statement of Purpose

A specialized questionnaire was prepared and sent to commercial data processing centers in the United States. The broad intent was to obtain a summary of the extent of scientific and technical data processing currently being performed in the various areas of science and technology, to identify what the major problems are in providing these services, and to probe the future role of such commercial data processing centers.

B. Survey Approach and Response

The sample was chosen from a population of approximately 1,000 data processing centers presented in the "Directory of Data Processing Service Centers" published in Systems Magazine, August 1967. Our sample of 422 centers resulted from the selection of those centers performing public data processing services. The centers chosen by this process were then screened by title in order to eliminate many of those specializing in business data processing.

Of 422 centers queried, 331 (about 78%) responded to the questionnaire. Many of the responses represented more than one data processing center, since questionnaires had been sent to all branches of several large companies, and most branches forwarded the form to their main offices to be completed. A total of 119 respondents, representing 225 individual centers, indicated that they do not process scientific and technical data; rather, they process business data. Approximately ten percent of these centers indicated that they would like to begin performing scientific and technical data processing in the future. An additional seven percent of those processing business data indicated a low volume of technical data processing, accounting for less than 5% of their total work load. None of the responses from this group were used in our final analyses.

Forty-five of the respondents, representing 106 individual data service centers, indicated that they engage in the processing of scientific and technical data to such an extent that it constitutes a significant portion of their business. The customers serviced by this group of centers represent all areas of science and technology. Tables IV-B-1 and IV-B-2 summarize the general activities of these respondents, and classify them according to the kind of data processed and the types of customers serviced.

Seventeen of the respondents service from one to fifty customers; thirteen service from fifty to 150 customers; seven service from 150 to 300 customers; and six indicated that they service over 300 customers. The remaining centers did not indicate the size of the community they serve. A reasonable estimate based on the above figures is that 6,500 scientific and technical firms use the commercial data processing centers surveyed.

It is difficult to determine the average number of data processing runs per week and the average length of time per run because of time-sharing, on-line operation, sporadic and diverse applications, and other related modes of operation. However, enough data were collected from centers to give an indication of the average number of runs per week. For respondents servicing less than fifty customers, the average number of runs is 396, while the average length of a run is 39 minutes. This group would average 257.4 machine hours of continuous operation each week. Respondents servicing between fifty and 150 customers reported an average of 2,785 runs per week, averaging 7.5 minutes for each run. The total for this group would be 348 machine hours per week. In the 150-plus customer category, centers averaged 2,446 runs per week, with the average run requiring 9.75 minutes. A total of 397.4 machine hours per week would be the average for this group. Extrapolation of these data indicates that the centers responding to the survey perform approximately 14,066 hours of scientific and technical data processing per week.

As part of the services offered, 87% of the respondents store customer data off-line. Although this was not unexpected, it was surprising to find that 53% of the respondents store and retrieve scientific and technical data as a major function of their data processing service. External modes of data storage mentioned by seventeen respondents included cards and tape decks. Internal storage of data was reported by five respondents. The predominant forms of internal storage are disks and drums.

TABLE IV-B-1

DISTRIBUTION OF CUSTOMERS ACCORDING TO TYPE OF ORGANIZATION

Organizations Served	No. of Respondents	
Government Agencies		17
Commercial Firms		67
■ General Manufacturers	16	
■ Aerospace Firms	11	
■ Private Enterprises (General)	11	
■ Civil Engineering Firms	6	
■ Electronic Manufacturers	5	
■ Oil Companies	5	
■ Chemical Companies	5	
■ Utility Companies	3	
■ Communication Companies	2	
■ Instrument Manufacturers	1	
■ Furniture Industry	1	
■ Pulp and Paper Manufacturers	1	
Universities		8
Professions		14
■ Consulting Engineers	10	
■ Oceanographers	4	
TOTAL		106

TABLE IV-B-2

DISTRIBUTION OF CUSTOMERS ACCORDING TO DISCIPLINE, TECHNOLOGY, AND WORK FUNCTION

Discipline or Technology	Number of Respondents
Civil Engineering	8
Mathematics and Statistics	6
Electronics	3
Aeronautical Engineering	2
Chemical Engineering	2
Nuclear Engineering	2
Social Sciences	1
Geophysics	1
TOTAL	25

Scientific and Engineering Function	Number of Respondents	
System Design and Development Processing		46
■ Engineering Design Calculations	16	
■ System Simulation	13	
■ System Development	9	
■ System Analyses	8	
Scientific Data Processing		22
■ Data Reduction	13	
■ Data Computations	9	
Manufacturing Data Processing		10
■ Technical Data Packaging	8	
■ Manufacturing Control	2	
Opinions Research		2
TOTAL		80

It was found that 70% of the service centers are independent, and are not linked in any way to a regional or national network. Those respondents which are components of larger systems represent a few sizable operations such as GE, NCR, COM-SHARE, Control Data, and UNIVAC. Several of the independent centers gave strong indications that they are in the process of converting to network operations, or that they plan to in the near future. As these centers continue to join processing networks, they may begin to play a more meaningful role in national data activities.

Most of the centers convey data by messenger to and from the customer. However, other means of communication are often employed, such as dataphones, TWX, telephone networks, over-the-counter service, mail, and telex systems. These channels of communcation are used to either support or supplant the messenger service. The needs of the customers determine the combination of services used.

Officials of data processing service centers demonstrate considerable insight into the complex problems they face at present and in the future. Unfortunately, most of the major problems they cite are the result of situations which they cannot control. Table IV-B-3 presents the most significant problems they reported, along with the suggested solutions.

The respondents identified several significant trends or changes which they expect to develop in the near future within data processing centers. The participants seem to agree that they will continue to have a healthy growth rate, and more business than they can handle. In addition, several specific observations were made:

- On-line, time-sharing, and the use of data links to remote terminals were rated as extremely important trends.

- Centers will become more specialized as the industry grows and as competition increases.

- A more professional attitude toward data handling and data handlers is expected to develop as the profession grows and matures.

TABLE IV-B-3

PROBLEMS ENCOUNTERED BY DATA PROCESSORS

Problem Statements	Suggested Resolutions
Personnel are not being adequately trained at a sufficient rate to fill the present and future needs of data processing centers. Constant equipment innovations and alterations, which necessitate retraining of personnel, compound the problem.	Comprehensive computer science and information processing curricula should be established at the undergraduate level. Seminars and night classes should be set up so that more management personnel can be attracted to data processing.
Customer understanding of the costs, limitations, and abilities of computers is inadequate.	Continued education of clients and potential clients by direct mail and public relations is necessary. Additional formal education on a broad scale is also required.
Continual changes in design of hardware and software cause many difficulties in data processing centers. Manufacturers have rushed to place many large-scale machines in the service bureau environment, thus causing serious difficulties concerning equipment compatibility and the choice of languages.	Additional time and effort should be expended in the design of hardware and software by manufacturers. New models should be introduced less frequently. Eventually centers must change from a time sales orientation to a scientific and technical services orientation. In this way, the emphasis can be shifted from the machine and language factors to specific applications and good, reliable service.

TABLE IV-B-3 (cont.)

Problem Statements	Suggested Resolutions
Common carrier transmission speeds and costs are restrictive. Tariffs, facilities, and policies need to be updated, for the systems of the future will require inexpensive and easy access to data processing centers from remote locations.	More direct attention should be given by the carrier companies to data transmission service, with a definite effort to reduce prices for large volume use, either by a single user or by a cooperative of small users. Competition for existing commercial carriers should be considered.
Data processing centers are concerned about the unfair competition currently given them by protected institutions--Government, universities and banks. These institutions do not rely on their services as a main source of income, and therefore they can offer noncompetitive prices.	These institutions should be subjected, in the computing service area, to surveillance by the anti-trust activity of the Federal Government.
Acquisition and dissemination of information about data processing programs are presently uncoordinated and inadequate.	A central library for inquiry as to where certain programs can be found, what individuals are specialists in what fields, and where to locate them, should be established.
A good set of standards in data collection and reduction is needed. Specifically, standards are needed for remote terminals.	A committee of experts to determine and monitor standards should be established. Under appropriate sponsorship, criteria should be established for the features of remote terminals that should be standardized.

TABLE IV-B-3 (cont.)

Problem Statements	Suggested Resolutions
A classification scheme or organizational framework that is common to all disciplines--a "common language"--is needed for scientific and technical data processing.	A non-profit organization, similar to the Ford Foundation, should be established to oversee the establishment of a basic scientific and technical information policy, to determine needs, and develop and implement the solutions of problems in scientific and technical data processing.

- More businesses are avoiding the rental cost of the computer by using the processing center services. This trend may be accelerated by high costs of programmers and the personnel shortage already in existence.

- Telecommunications will become an increasingly common communication vehicle.

During the preparation of this report, considerable insight was acquired into the workings of the commercial scientific and technical data service center. At the present time, centers are operating near total capacity. The outlook for the next ten years, in terms of growth and extended, sophisticated services, appears healthy. On the other hand, the community is faced with a group of complex issues which must be resolved in order for the industry to maintain its present position and realize its expected growth. It is, therefore, suggested that these problems be given closer study and that actions be taken to find satisfactory solutions to the problems. If these problems could be solved, the role of data processing centers would be immeasurably strengthened, and corresponding improvements would take place in data activities and systems affected by the centers. For example, greater customer protection against unfair practices, encroachment upon individual rights, and other illegitimate activities could be effected by standardizing and enforcing ethical practices and procedures within the industry.

V. SUMMARY OF SCIENTIFIC AND TECHNICAL DATA FILES OF THE DEPARTMENT OF THE ARMY

-- A Selected Inventorying Probe of Existing Data Resources

Discussions with leading data management specialists revealed that large volumes of scientific and technical data are generated in pursuit of mission objectives. Except in limited cases where analyses have been conducted preparatory to design and introduction of computer or other modern data handling methods, few inventories have been conducted of these types of data collections. Performance of extensive inventories was beyond the scope of the current survey; however, it appeared desirable to obtain and present typical summary information concerning the data inventory within a selected mission area.

Fortunately, in 1966 the Howard Research Corporation, as part of Task I under the Department of Army Engineering Data and Information System Development Project, published a review of all recent inventory activity covering scientific and technical information and data within the Department of Army Research, Development, Test and Evaluation Programs.* Information was compiled which described, in a directory format, the scientific and technical information holdings of 171 organizational units. Table V-1 summarizes the volume of data-documents or data artifacts held by these units of the Department of the Army. A total of 32 different forms of data were identified.

This summary of the Army's mission-oriented data collection merely serves as a sample of the many collections of its type which have not yet been inventoried.

---

*EDIS Task I Report - Categorization of Existent Data Systems, Howard Research Company, Division of Control Data Corporation, 20 January 1966.

TABLE V-1

SUMMARY OF DATA COLLECTION

Form of Data	No. of Systems Reporting	Number Reporting Volume	Total Volume
Medical Case Records	3	3	76,480
Microphotographic Reels	2	2	1,181
Microphotographic Strips	3	2	7,100
Microscopic Slides	3	2	18,000,050
Patents and Patent Applications	14	9	55,283
Photographic Chips	1	1	300
Photographic Negatives	35	29	600,497
Photographic Reels	15	9	4,430
Photographic Strips	7	4	265
Photographs	42	30	145,386
Punched Cards	22	18	1,505,060
Punched Cards, Edge-notched	5	4	3,951
Punched Paper Tape	3	3	42,000
Silent Motion Pictures	25	14	1,511,709
Slides	45	35	94,296
Sound Motion Pictures	25	16	312,839
Specifications and Standards	48	36	81,039
Tissue Specimen Slides	3	3	1,147,100
Video Tapes	2	1	2
X-Ray Films	3	3	99,000

VI. REVIEW OF CURRENT EQUIPMENT CAPABILITIES FOR SCIENTIFIC AND TECHNICAL DATA HANDLING

A. Statement of Purpose

Prerequisite to development of a time-phased plan for the development of national data systems and programs is a knowledge of the data handling capabilities afforded by presently available hardware, and the developments which will influence these capabilities in the future. This report section summarizes the present state of the art in data handling equipment and the significant development trends in this field. It presents the findings resulting from an extensive review of the current literature, and reflects observation of usage in science and technology, as illustrated in Table VI-A-1.

B. Reporting Structure

The approach used in reporting the findings of this review of data system capabilities is to follow the pattern of data flow. Accordingly, the first equipment capabilities to be discussed are those associated with the input of data into systems. The second topic is storage and retrieval equipment, and the next is output equipment. Then, those equipment capabilities associated with interactive input and output, and those concerned with transmission are treated as separate topics. Finally, in the last section, a summary is presented highlighting some perspectives of present and future equipment capabilities which constitute application and development problems.

C. Review Findings

1. Input Equipment

Keypunching is still the most widely used input technique. Other developments are appearing, however, and will be discussed. Their usage lies mainly in special-case and unique applications.

TABLE VI-A-1

TYPICAL USAGES OF COMPUTER SYSTEMS IN SCIENTIFIC AND TECHNICAL DATA ACTIVITIES

Field	Application
Aerospace engineering	Data documents (drawings, reports, specifications) for large projects are centrally stored and retrieved using computers.
Space science	Microminiaturization of data storage and processing equipment for on-board satellite use and data telemetry.
Structural engineering	Use of computerized data bases for breakdown of large structural concepts into basic elements to facilitate component design.
Chemistry	Thermodynamic data are compiled, evaluated and disseminated on magnetic tapes for use in computer calculations of rocket fuel performance, etc.
Biomedical science	Computerization of clinical data for teaching and diagnostic use.
Social sciences	Use of networks of computer systems to store and retrieve large volumes of survey data for social research purposes.
Oceanography	Storage of bathythermographic data in computer memory and automatic extrapolation of data for areas for which there are no measurements.
Chemical engineering	Computer system used in textile plants to automate matching of millions of dye colors for textile manufacture.

New optical character readers are gradually coming into wider use. However, they still face the problems of variability and lack of standardization of input material. Another substitute for keypunching is offered by the magnetic tape encoder, but these are unable to insert material at random. Their application is thus restricted to general-purpose usage.

In some cases, incremental magnetic tape encoders do provide a time and money saving advantage in comparison to card-producing keypunches. With these encoders, keystroked information is entered directly onto computer-compatible magnetic tape. This bypasses the time-consuming car-to-tape conversion or a direct, slow card input to the computer. These encoders offer the advantage of easily verifying, or transmitting, information from one tape encoder to another over telephone lines. Another advantage is that keystroke errors are reduced because each record is stored in an 80-character buffer memory and is open for visual verification before entry onto tape. This system does not, however, allow easy editing of the entire keystroked file as does a punched card deck, because of the serial nature of magnetic tape. Cathode-ray-tube (CRT) display consoles can also be used for direct entry into the computer of programs and other data with instantaneous verification and correction in the input process. An example of such use is data display in the Deep Space Network operations room at Jet Propulsion Laboratory.

A system that eliminates keystroking altogether by automatically scanning and reading printed or written source data holds the most promise. Development of such optical character readers, with the ability to read a variety of source data, has been relatively slow. Mainly, they are still limited to reading highly formatted material. Only one or two are capable of reading handprinted data, and there is disagreement about the ultimate need for such equipment in science and technology. Cursive script readers are not as fully developed as print readers. Two of the problem encountered are the difficulty in determining where one letter ends and the next begins, and the great variety in cursive scripts.

Another problem that limits the use of optical character readers is the requirement of paper handling. Speed increases can be obtained by multiple scanning of documents or by multiplexing the input. The

paper handling equipment for these systems is expensive, often costing two to three times that of the basic reader. In spite of some claims that optical readers will replace keypunches for nonhandwritten source data in the next few years, there is still no clear indication of standardization of reader fonts and papers. It remains to be seen whether or not such standardization will occur.

2. Data Storage Equipment

(a) Digital Storage. The trend in bulk scientific and technical data storage is to utilize digital computers and storage devices that are smaller and faster even though they are more complex. To give computers increased read-only storage, various techniques using optical and film technology are being used. The combination of film with the high resolution and precision of laser and electron beams is being used to increase the information capacity of film.

In the area of erasable input/output memories, large direct-access magnetic memory devices are being developed. Though more expensive than standard magnetic tape units, they are being used more and more because positioning time is extremely fast. One of the more important trends in this field is toward removable-medium devices. Like tapes, these allow almost unlimited storage, since the information-bearing magnetic surfaces can easily be replaced. Information is recorded on magnetic strips or cards, tape loop cartridges, or disc packs.

(b) Internal Computer Memories. The trend in the area of internal computer memories is toward smaller sizes and faster readout speeds. Today, there are memory cores with capacities ranging from eight to thirty-two million bits, with cycle times of three-fourths of a microsecond. That compares with the ones of the early 1960's that had a capacity of about one million bits with a cycle of two microseconds. Not only have the capacities and speeds increased, but costs have reduced on the average by a factor of four.

At the same time that high speed, large capacity memory cores have been developed, so also have even larger capacity, but slower, storage ones been perfected. These are basically an extension of internal memory that can be attached on the new computer systems.

This increase in core capacity allows the use of large programs and data bases (such as those in the space sciences) more effectively than before. For example, data from a program can be held in slow internal memory, saving on positioning and latency, then moved into fast memory when required.

Magnetic cores remain the most widely used main memory elements, with research and development trends toward smaller, faster units that allow for more compact memory units to be designed. IBM, for example, has developed cores with outer diameter sizes of only 0.0075 inches--small enough to fit inside the center hole of more cores.

Beginning to enter the memory core field are new memory elements such as plated wires, planar thin films, monolithic ferrites, and integrated circuits. Because these new units have lower production costs (since they are being batch-fabricated in one step), they are expected to become competitive quickly. They also have a speed advantage, already taking over in the fields of high-speed registers and temporary "scratch-pad" memories. The superconductive cryogenic techniques that once were potentially useful for quick, on-line storage have been held back because of the high costs of refrigeration.

(c) Content-Addressable Memories. The development of content-addressable memories, to facilitate the location of information by content instead of address, presents both promise and limitations. Potentially, the technique holds promise on somewhat the same order as retrieval by document content, compared with retrieval by accession number. The problem, so far, has been the relatively high cost of the electronics required for each cell of the memory. Widespread use of integrated circuitry, and the companion price reductions, will reduce the cost of this equipment and should bring increased development. The inherent limitation is that for the content-addressable memories (CAM) technique to be successful, it must be supported by the capabilities of a general-purpose computer.

(d) Read-Only Storage for Microprogramming. A read-only storage technique has been developed as a new way of organizing the execution machinery of a central processing unit. It uses lower-level commands to encode or microprogram machine-language instructions, rather than providing circuits to execute each machine-language instruction. This means that two separate circuits are built into the machine; one for handling the lower-level commands, and the other for translating a machine-language instruction into its microprogram. Most computers using this system of organization have the instruction set defined at the factory. In some cases, the instructions are decoded using a read-only storage. Multiple read-only systems are also available that can emulate older machines. This saves rewriting existing programs, particularly when run on the newer, faster computers. Microprogramming techniques are expected to expand with many forms of new instructions being devised.

(e) Large-Scale Integration. A promising development to improve computer digital storage capability is the rapidly growing large-scale circuit integration technology. This involves building complex circuit functions into tiny chip semiconductor material using such techniques as micro-etching, micro-plating, and micro-evaporation. These techniques enable more compact and complex integrated circuitry to be built more cheaply and more reliably than older forms of integrated or discrete circuits.

Cost reductions come through the ability to batch-fabricate such systems. Conversely, batch-fabrication is vulnerable to having a production defect reproduced in large numbers, necessitating the discard of numerous components. Techniques to prevent this are being developed; such as discretionary wiring. Tests are made for defective cells in a redundantly constructed integrated array with the good cells selected out. This technique of integrating circuits also has application to memory construction. Eventually, it may be possible to produce both the comparison circuitry and memory cells of content-addressable memory into a single unit.

(f) Image Storage and Retrieval. One of the most significant advances in the storage and retrieval of data documents is the improved accessibility of images on microform resulting from new

equipment for producing, retrieving, reading, and printing images. Of special significance has been the development of new facsimile equipment for retrieval and long distance transmission.

Some typical equipment will be discussed, first in the smaller sizes, then on a larger scale. The Houston Fearless Corporation has a desk-top microform retrieval unit and reader with any of 73,500 pages (750 fiche with 98 images each) accessible in less than 5 seconds. Costing in the $5,000 range, the system has an optional hardcopy printer or a teletype or CRT for on-line hookups. It has edge-notched identification numbers stored on the metal frame surrounding the microfiche. The search request for a random page, or next page, is entered via keyboard, paper tape, or computer to maintain file integrity.

The Mosler Safe Company's Selectriever is another similar system. It provides for 6-second needle-sort retrieval of any one of up to 200,000 aperture, microfiche, or tab cards, which are stored in 100-card cartridges. The cost of this unit is in the $40,000 range. It can be interfaced with a computer, hardcopy printer, remote displays, or facsimile transmission system.

Examples of larger systems that cost in the $1 million range are Magnavox "Magnavue" and the IBM Cypress 1350. The U.S. Army Missile Command uses the Magnavue System in its Documentation Automated Retrieval Equipment (DARE) program for the processing of large file engineering drawings and associated documentation. This is a computer-controlled system that provides automatic collection, storage, retrieval, and preparation of punched Diazo copy card outputs. This is done from a rapid-access file with a capacity of 750,000 microfilm images. In random mode, the equipment provides an average access time of 50 seconds. In normal sequential batch processing, a file of 750,000 microfilm images can be processed in about seven hours.

A typical daily sequential processing includes the output of about 5,000 Diazo copy cards, the input of 2,000 new Magnavue film chips, and the removal of 1,800 Magnavue film chips from the system. The film chips are 33mm wide by three inches long on a Mylar base. They are held in 2,500-chip magazines. The film chip

has a high-resolution image portion for the storage of a standard 35mm microfilm image and a coded portion for the storage of 80 alphanumeric characters that identify the associate microfilm image. It can also use magnetic chips of the same size to store additional information associated with a given document.

The IBM Cypress system has storage facilities for 500,000 images on Mylar-based chips. To retrieve, a cell of 35mm by 70mm chips is located under computer control and transposed pneumatically to a copying station. Here, the selected chips are copied onto an aperture card. This is then transported for later viewing or reproduction. The total retrieval time for this type of random access system is from four to six seconds.

The two systems described above are not capable of being updated by erasure. Instead, a physical substitution of records must be made. Ampex's Videofile system is particularly suited for real-time retrieval of documents that have a very short life span. Indexing is done numerically or alphabetically, using 18 numeric or 12 alphanumeric characters stored contiguously addressed to images on video tape. Access to a document can be remote, with a hard or softcopy output. Storage runs about one page per one-third linear inch of standard two-inch wide tape.

In the immediate future, it seems likely that ultra-high linear reduction ratios on the order of 150:1 to 300:1 will be achieved. This compares with the standard 15:1 to 25:1 of today. This would be extremely valuable in solving the bulk storage and retrieval problems, allowing for the dissemination of copies cheaply and providing real-time access to large million-page files.

The IBM "trillion-bit" 1,350 storage device is a good example. It uses 35mm x 70mm silver halide film chips. A total of 4.5 million bits are prerecorded on each chip by an electron beam. For read-out, a plastic cell containing 32 film chips is sent to a selector. This picks the proper chip from the 32 in an average access time of six seconds. After the chip has been positioned, information is read using a flying spot CRT scanner.

The Precision Instrument Company's UNICON system is another approach to storing digital information optically. This system uses a laser to write 0.7-micron-diameter holes in the pigment of a film. The information is organized in records of, at most, a million bits, with each record in a 4-micron track extending about a meter along the film. Each record is identified by information stored next to the beginning of that record, in an additional track for the proper code, then scanning the track with a laser weaker than that used for writing. Predictions are that one UNICON device with 35mm film could store a trillion bits on 528 feet of film, with an average access time of 13 seconds.

While optical storage systems provide substantially higher information densities than are achievable with solid-state or electromagnetic memories, their non-erasable nature makes them more suitable for storing relatively permanent reference data such as thermophysical properties, than operational data such as those used in chemical process design. However, the ability to have computer access to such a vast amount of storage means that radically new and more efficient approaches can be used for information storage and retrieval installations.

3. Output Equipment

Far more progress is being made in the development of output equipment than input devices. Basically, it is much simpler to convert machine-sensible information to a human-readable form than to accomplish the reverse operation. Special high-speed printer chains with both upper and lower case letters are now used for high volume printing. Impact printers are already well advanced.

Advances are being made in nonimpact printers by adapting long distance xerography to computer input/output. This takes bit-by-bit scan information and converts it to graphical information on standard paper at a maximum rate of 768 lines per minute. Advantages of this output system are that no limits exist on character set and graphical information can be printed directly. The disadvantage is that the system is slower and more expensive than high-speed impact printers.

A significant trend in output technology is the use of microform for computer output. This reduces handling costs, saves in printing time of master and dissemination copies, and in physical volume. Increased demand for microform printers is prompting manufacturers to go into second generation equipment with increased capabilities. Micro film recorders convert digital codes from a computer to their alphanumeric and graphical forms on a cathode ray tube (CRT), where it is recorded by cameras. Some of the larger units convert alphanumeric computer output into microfilm at the rate of 9,000 pages per hour. Hardcopy equipment can be used with the recorders.

A new trend in this equipment area is the elimination of the traditional wet-stage development used with conventional microfilm by using a new dry method. Output from these new recorders is written directly on the microfilm with an electron beam at a rate of about 30,000 characters per second. This bypasses the filming of a CRT data display and yields a sharper picture, in addition to real-time, high-speed output.

The advantages of microform computer output suggest that all output, allowing for special exceptions, may eventually be converted directly to some type of microform, eliminating the intermediate step of hardcopy. Additional development is required, particularly in the microform reader/printer technology, but the potential for this approach is sufficiently high to motivate continued development effort.

4. Interactive Input/Output

For some years, teletypewriters have been used to provide real-time data communication with a computer. This is most common with small computers or in time-sharing applications. There is growing use for this type of application in scientific information retrieval, text and message manipulation, and problem-solving for high-speed interaction, or "softcopy." Some approaches to this capability are discussed.

(a) Audio Input/Output. Limited progress has been made in the transmission of information (i. e., data requests) to people from

a computer via telephone. At one aerospace systems company some 14, 000 engineers have telephone access to a design data bank which audio-transmits design criteria and specifications. Computers "speak," using voice response units. Typical equipment in this field is the IBM 7770 Audio Response Unit, the IBM 7772, and the Cognitronic's "Speechmaker."

In the case of the IBM 7770 Audio Response Unit, the system has a vocabulary of about 128 phrases and can accommodate some 48 telephone lines simultaneously. The phrases are magnetically stored as an audio signal of one-half second duration. Messages consist of a playback of a sequence of recorded phrases with the computer controlling the message selection, switching, and sequencing.

The Cognitronic "Speechmaker" audio response units use a prerecorded word technique that is somewhat akin to the way a sound track is applied to motion picture film. The American Stock Exchange is one example of the use of this system. Brokers can dial four-digit codes from their office telephones and get stock quotations with approximately 1, 200 inquiries handled per minute.

IBM's 7772 system stores words in digital form, with the computer generating commands for a speech synthesizer. Since digitally coded phrases can be of any length, a flexible vocabulary is available. Limitations are whatever computer memory is available. An extension of the computer audio response is under experimental development by MIT's Research Laboratory of Electronics. The goal is to perfect a reading system for the blind, using computers. A three-part program is under way. This involves character recognition, translating words into the minimum units of speech, then converting this into speech, using speech synthesizers. Progress has been made, though much work remains.

In this same technical area are audio couplers. These are acoustical devices that allow telephones to input audio signals into computers in the form of digital data. Tymshare's Audio Magnetic Data Transceiver, for example, takes digital signals, converts these to acoustical signals that are transmitted over the telephone line to a conventional data printout set at a computer terminal.

This equipment is particularly useful with portable teletype equipment, putting a computer terminal at the user's fingertips.

The area of voice input to a computer lags behind output techniques and must be considered as still in the research and development stage. Limited progress has been made, but so far, no operational speech recognition systems have been perfected.

(b) Video Input/Output. Video interaction with computers using softcopy displays is widely in use and becoming increasingly important. This contrasts with the field of audio interaction which is still largely in the research stage. Sufficient experience has been obtained in video interactive equipment to establish its cost-effectiveness for a wide variety of uses. It is widely used, for example, for remote-access stock market quotations, and certain aspects of computer-aided design and documentation in the aerospace, automobile, and computer industries. In the latter uses, the two-dimensional feature of the cathode ray tube provides far more flexibility for sketching and plotting than is possible on printers or teletypewriters.

Display consoles are compatible with third generation computers, adding to their acceptance. Time-sharing and multiprogramming techniques have advanced to the point where displays no longer need monopolize a central computer's time for applications which demand the ability to indicate alternative choices with a pen or cursoe, which adds to their flexibility. Generally, displays fall into four basic categories: TV monitor output-only consoles; large-screen group displays; alphanumeric input/output consoles; and alphanumeric plus graphic input/output consoles. This latter category is commonly used in computer-aided design work because of the wide variety of input devices available such as the light pen, alphanumeric keyboard, the Rand tablet, and the like. The RAND Tablet, a graphical man-machine communications device, is potentially one of the more fruitful approaches for two-dimensional graphic input to a computer. The high resolution of the tablet, high data transfer rates, and ease or "naturalness" of use are its chief assets. These same characteristics, however, give rise to the major problems in designing the tablet/computer interface. These problems are amplified when the interface is required for a multi-terminal, time-shared computer. However, recent developments aimed at solving these problems are promising.

For the typical time-sharing system user's console--a keyboard/printer--to be considered interactive, it must receive some response from the computer in fractions of a second, and a certain amount of computational response within several seconds. With the graphic tablet display console, however, the computer reaction time must be on the order of a few milliseconds, and computational response, on the order of one second. The high data transfer rates and the speed of response required for user psychological reasons demand that the user-to-system interface be "tightly coupled."

The trend in modern computer system design is to provide external input/output buffers for cathode ray tube and graphic consoles; thus, data are not readily accessible to the user program. Typical systems require block transfers of all the data from the console into main core for processing, maintenance of the complete image of display tables within the user's core space, and block transfers back to the input/output device.

Even though the improved techniques in time-sharing and multiprogramming are reducing the operating costs of the central computer, prices are still relatively high for some of the associated display equipment. The trend is to reduce the cost of this equipment as more comes into use and new technical advances are made. General-purpose displays and, to some extent, alphanumeric display consoles, can give a wide range of file organization, but they are also relatively expensive. Such alphanumeric softcopy displays have less capabilities than the general-purpose equipment, but also cost less. They usually have a single keyboard and cursor input, but no light pen. Displays are connected to the computer through a multidisplay control unit that contains the display logic, buffer storage, and possibly local message editing and formatting capability. This keeps the display independent of the computer except for short bursts of messages.

This arrangement has the advantage over standard teletype terminals of speed and silent operations. Their control units accommodate teletype or other forms of hardcopy output that can be initiated from the display. The trend is toward development of new input devices and display media. Such equipment, for example, as IBM's 2260 and Raytheon's DIDS-400 seem certain to have wide application as time-sharing terminals, information retrieval systems, and other such uses.

For the immediate future, it seems that cathode ray tubes and direct view storage tubes will continue to dominate the field. Work is being advanced on digitally driven panel displays with built-in memory.

5. Image Transmission

A key factor for any data or document retrieval and dissemination system is the ability to transmit resource material from central locations to remote stations. It is now possible to link transmission devices directly to microfilm retrieval systems and provide a hardcopy alternative to closed-circuit television. One such system is the Alden/Miracode system that integrates Alden's Alpur-Fax facsimile system with Kodak's Miracode automated microfilm retrieval system. It scans documents in the microfilm viewer and transmits the information over telephone communication lines to make available hardcopy at remote locations. This is done at the rate of three minutes per page.

Another such system is the Xerox Magnafax Telecopier. This is a facsimile device using normal telephones for transmission. An acoustic coupling mechanism makes the system portable. Its copier is a continuous-scanning facsimile transceiver. Photocells pick up reflected light from the document being transmitted. This light is converted into frequency-modulated audio that is transmitted over the telephone. At the receiving end, the document is reproduced by two mechanical styli on special carbon backed paper.

A restricting factor in widespread use of image transmission systems at present is the relatively high cost involved. Line charges are one of the major cost factors because usage normally comes during prime telephone rate periods. Though it is becoming possible to transmit information on a real-time basis from central locations, the cost involved restricts this at present to high-priority usage.

D. Summary of Findings

Four major trends dominate the present activity in the development of improved data storage, retrieval and dissemination capabilities:

- There is a trend toward the development and use of smaller, relatively inexpensive equipments to meet the increasing requirements of less expansive data systems such as clinical data systems being installed in most hospitals;
- There is a trend toward the expansion of storage and speed capabilities of central storage units of large data systems, such as those involved in collection and networking of data associated with weather, and other gross phenomena (see Table VI-D-1);
- To facilitate networking of data efforts, there is a trend toward improved remote consoles, use of multiprogramming, and satelliting of small computers connected to central storage and processing units; and
- To satisfy demand for improved data input and output as well as interactive input/output equipment, there is a trend toward faster CRT, faster response times in the order of 10 microseconds, and a new generation of display equipment that may even provide three-dimensional display capability.

The following pages elaborate on these four areas of development, presenting first the trends and problems in computer development; secondly, the trends toward networking of data efforts and the associated equipment problems; and thirdly, the underlying problem related to equipment development and utilization.

1. Computer Systems Development

There are two significant trends in computer development. One is toward large, complex time-sharing computers. The other is toward decentralization of computing power through small computers. Each system has advantages, and unique areas of application. While time-sharing is growing rapidly and offers many advantages in a large number of situations, small computers have sufficient attraction in their own right to insure that they, too, will continue in use.

The practice of time-sharing on computers has grown rapidly in the past few years, starting with the general-purpose computers and

TABLE VI-D-1

REPRESENTATIVE COMPUTER CHARACTERISTICS AND COST 1950-1975*

Year	1950	1955	1960	1965	1970	1975
Add Time (not including access time)	240 μ +	15 μ +	4 μ	.8 μ	8 nsec‡	1 nsec
Memory Cycle	282 μ	12 μ	4 μ	.5 μ	100 nsec+	30 nsec
Storage Bits	1000 delay	1.5×10^6	1.5×10^6 +	6×10^6	12×10^6	$10^8 - 10^{11}$
$ Cost/Bit Core Storage	2.61	1.78	.85 +	.20	.10 to .3	.005
¢ Cost/Bit Magnetic Recording Storage	.3 to .7	.1 to .5	.5 to .001	.1 to .001	.001 to .0001	10^5 to 10^6
$ Cost/10,000 Instructions (Average)	.10 to .15	.01 to .05	.005 to .01	.002 to .0005	.0005 to .0001	.00001

No definite source is available for exact costs. They are estimated from rental and size of computer. All these figures are the general representation of each era of computer, and not a specific product. Symbol * near u is for microsecond (one millionth of a second). Symbol ‡ near nsec indicates that the term stands for nanosecond (one billionth of a second).

*Source: "85,000 Computers by 1975, Operation Cost to Drop, Comprehensive Report Reveals," Administrative Management, June 1966, p. 53.

increasing to the point where special time-sharing equipment is now being built. More major companies in the computer industry have new time-sharing computers either developed or in some stage of development.

Most of the new systems in time-sharing center around the technique of multiprogramming, or multiaccess programming, where several problem programs are basic in the main computer memory. This permits the supervisor program, in the main memory, to interleave execution of different problem programs, keeping the time waste for completion of input-output operations at a minimum.

To protect against a defective program damaging the supervisory program or other programs in the memory system in multiprogramming, protection is provided in the time-sharing hardware. In addition to protecting against defective programs, the safeguards also provide for privacy.

To achieve the fullest benefits from multiprogramming, the maximum number of jobs must be kept executing at the same time. The limitation on this is the availability of main memory capacity. In some cases, enlarging the memory capacity increases computer efficiency. Another technique is to keep only part of each job in the main memory with the balance held in a high speed mass memory device such as a drum. Parts of programs are then swapped in and out of the main memory system. To aid in such swapping, memory paging is used, with blocks of the main memory assigned addresses in a wide range. With this technique, the supervisor program can assign problem programs within this range.

Parallelism is another important organizational feature of large computers. It separates the central processing unit from the input/output devices with a subordinate processor. The technique has been perfected to the point where five to ten arithmetic instructions can be performed simultaneously within a central processing unit while still more data are being transferred from other slower memory units. Parallelism has opened the way to multiprocessing, with two or more independent central processing units sharing some facilities such as input/output devices and several banks of the main memory. This is important in time-sharing, since part of

the system can be out of service without nullifying the whole system. The key to time-sharing facilities is a complex supervisor program that utilizes their fullest potential.

Accompanying the development of the extremely large, complex computing systems, there is, similarly, a trend to perfect the very small computer. These meet the needs of users where only limited requirements exist. A wide range of optional equipment is available to extend their capabilities. Users, however, need only add the equipment they need, and thus are saved from acquiring extra capacities that are surplus to their needs.

The trend in small computer development is toward miniaturization, with some measuring only a few inches in size. Another trend is to standardize coding of characters to 8 bits. Still a drawback to computer-sharing is the communications cost of leasing the telephone or teletype lines between the user and the computer. Often, the cost of using the computer is little more than the cost of the communications link to reach it. This communications cost remains constant, regardless of whether full or partial utilization is being made of it.

What appears to be needed is a communication system which will allow the user to communicate with the remote computer as the need arises--with a minimum of effort and delay to the user and at a price substantially lower than that of the remote computer itself. Under the existing service and rate structure, the user of a remote computer utility is faced with several alternative methods of using the communication channels.

First, he may maintain the connection between his console and the computer for the duration of his computing session, even if, for the majority of the time, the communications facility is lying idle. Or second, it he wants to economize, he may break and, subsequently, re-establish the connection for only as long as it will actually be needed.

This second alternative has immediate economic advantages, but would prove cumbersome in practice. Furthermore, this break-and-reconnect approach presents a more complex problem in advance planning. Delays are likely, too, that would tend to decrease the

overall response time of the system and make true conversational data retrieval and computing more difficult.

Thus, while time-sharing systems are designed to provide the computer user with the opportunity to work at his most advantageous speed and interact with the computer at his convenience, available communications services have not as yet been designed for efficient and economical time-sharing computer usage.

Plans have been suggested which would divide communications facilities among many users; each user accessing the facility for brief periods of time. Present technology would allow a group of users to construct a shared-carrier operation by leasing conventional circuits from the common carriers. But, while this is technically possible, the trend appears to be to have the communications common carriers take the initiative and offer a sharing service. Charges for communications would be based on the amount of information transmitted, rather than the time the circuit was open. Irrespective of which approach is eventually taken, it seems clear that, unless large monolithic systems are implemented, the full economic advantages of time-sharing cannot be attained.

2. Data System Networking

The most significant trend in data processing equipment development is the evolution of large systems and networks centered around families of equipment at one point instead of piecemeal, adding them on. This will increasingly make it simpler for the user to meet individual demands over a wider range of requirements.

Two approaches are being taken for inter-system developments. One is to standardize the system so that information storage and retrieval programs written for one system can be run on another. The other approach is to attempt to satisfy the constant requirement for faster, larger, and less expensive main memories. Though there has been some slowdown in the requirement for higher speed memories, the demand for larger memories continues to grow and can be expected to do so. This need for larger memories has contributed to development of highly modularized systems. This trend is expected to hold

through the development of third-generation equipment, if not beyond this point. In turn, the highly modular system, because of its inherent ability to process more than one job at a time, will feed back to the memory area an ever increasing demand for more memory.

Interwoven into the memory/modularity development is yet another trend. That is, as the software developments gain more of the goals to be achieved, there will be a blending of the hardware/software systems. More hardware specifically designed to aid the software will evolve and a more efficient and intelligent use of the hardware will be made by the designers of customized software.

Accompanying these are a series of other developments that potentially could bring revolutionizing effects in the area of handling data. These include development of remote terminals, improved man-machine interfaces, and the communications links between central storage and processing units. As the central processor becomes more and more powerful and the software and hardware provide a networking capability, the "time-sharing" and the multiple user features of the systems will increase. There are still major hurdles to cross, however. Improved terminals are required to provide for practical remote data input and output. In addition, improved linkage of satellite processing units is required. These developments imply coordination of activities in the communications industry, the computer industry, and the user communities. The ultimate success will depend upon the ability of all to provide systems adequate to handle user requirements at viable costs. Specific recommendations relevant to equipment developments are included in Volume I of this report.

3. The Underlying Problem

This survey revealed that there has been relatively little communication at the planning level between the computer industries and the user communities, and that there has been little differentiation between systems for storage and retrieval of reference materials that contain data. Moreover, the data processing capabilities originally developed for the storage and retrieval of business data, and those developed for scientific and technical computations, have not been viewed to any large extent as easily applicable for storage

and retrieval of scientific and technical data. Therefore, most scientific and technical information systems in existence today provide access to reference materials such as documents or maps, rather than to the contained data, and the available capabilities for storage and retrieval of scientific and technical data have remained largely untapped. Two underlying causes exist for this situation:

(1) The traditional modes of scientific and technical data flow are based on the use of artifacts such as documents, and

(2) Many activities associated with storage and retrieval of scientific and technical data involve intellectual processes which seem too complex to economically program except in instances where extensively repetitive operations are involved.

Therefore, among the steps which must be taken to facilitate the utilization of existing and developing data system capabilities are coordinated implementation of certain equipment developments, training programs, and national programs. These are the subject of Volume I of this report.

The data processing equipment industry has passed through three major generations in equipment development and utilization. Each has been predominantly oriented to the component used in the logical portion of the central processor. The first generation of computers, primarily developed to satisfy the requirements for large-scale computation and business data management, was dependent upon machine-coded programs and used vacuum tubes. Early in the 1950's, assembly-type programming systems and transistorized circuits came into wide usage. In the 1960's, FORTRAN (Formula Translation) and COBOL (Common Business Oriented Language) had become standard practice, and integrated circuitry came into widespread use in a third generation of computer systems. But, while the basic computer hardware is moving in this third generation of development, much of the peripheral equipment is still in the first and second generation stage, and these equipments are proving adequate for the needs of scientific and technical data processing.

The trend toward remote terminals is also influencing the development of data processing equipment. For this reason, there is now a push in the perhipheral equipment field to "catch up." Each year, industry comes out with a long list of new gains or promising developments in this field. As this equipment moves into a new generation of development, certain trends are coming. Principally, there is a move toward making the new terminal and peripheral equipment electronic, rather than mechanical or electromechanical. Secondly, many are designed specifically to act as a transducer between the machine and the man.

Two problems have arisen with the growth of peripheral equipment. Both pertain to communications. One is the actual problem of communications channels, their cost and availability; the second is the interface with the computer. Originally, the interface was handled through the central processor. Now, it is being re-oriented to feed directly into the main memory. Both problems are being solved, though not always as rapidly as some users would like.

The trend in hardware development for equipment to handle data is to use it in a paired concept and sometimes in more complex arrangements. Such things, for example, as linking computer output directly onto microform, or an arrangement for computer-controlled microform retrieval, the linking of hardcopy and micro-form transmission over microwave and telephone lines, or the display-centered browsing of both image and digital files.

There seems little doubt that the technological advances in hardware to handle scientific and technical data will rapidly outpace the software aspects, making these the limiting factor in how fast progress is achieved. Coupled with the technological advances is the companion reduction in cost and availability. Most information systems in the near future, for example, are almost certain to have increased access to computer power either through local, small computers or on a time-sharing basis.

The old line concept of computers and their usage is rapidly being revised. Today, "computing" encompasses a broad spectrum of functions well beyond the traditional arithmetical computations. In fact, one of the fastest growing usages in the computer area is that